AN INVESTIGATION OF THE EFFECTS OF CORRELATION AND AUTOCORRELATION IN CLASSIFIER FUSION WITH NON-DECLARATIONS

THESIS

Frank M Mindrup, Captain, USAF

AFIT/GOR/ENS/05-13

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GOR/ENS/05-13

AN INVESTIGATION OF THE EFFECTS OF CORRELATION AND AUTOCORRELATION IN CLASSIFIER FUSION WITH NON-DECLARATIONS

THESIS

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Operations Research

Frank M Mindrup, BS

Captain, USAF

March 2005

AFIT/GOR/ENS/05-13

AN INVESTIGATION OF THE EFFECTS OF CORRELATION AND AUTOCORRELATION IN CLASSIFIER FUSION WITH NON-DECLARATIONS

Frank M Mindrup, BS
Captain, USAF

Approved:

//signed//	11 Mar 05
Kenneth W. Bauer (Chairman)	date
//signed//	16 Mar 05
Lt Col Robert T. Brigantic (Member)	date
//signed//	15 Mar 05
Mark E. Oxley (Member)	date

Abstract

Air Force doctrine requires reliable and accurate information when striking targets. Further, this doctrine states that fusion should be utilized whenever possible to ensure the best possible information is conveyed; there is no specific guidance as to how to fuse this information. This thesis extends the research found in Leap, Bauer, and Oxley (2004) to include a non-declared class. The Identification system operating characteristic (ISOC) was adapted to allow for non-declarations both at the individual sensor level as well as the fused output level. A probabilistic neural network (PNN) was also used as a fusion technique. A cost function was developed that incorporated misclassification error as well as non-declaration rules. In addition, a heuristic was developed to find optimal rules through a likelihood ratio method. Finally, a sensitivity analysis was performed.

To my wife

Acknowledgments

There are so many people deserving thanks for this effort. First off, I would like to offer my sincerest gratitude to my faculty advisor, Dr. Kenneth Bauer, for his guidance. Thank you for helping me to see the end during all of those times when it seemed out of reach.

I must admit that I was standing on the shoulders of giants for this effort. I want to state my sincere appreciation to Nate, Trevor and Tim for their tremendous assistance. I would also like to thank Clint, Tim, and Kevin for the comedic relief and motivation. I would also like to send a shout out to Jeff, Casey and JD for the study sessions and study breaks.

For my final and most heartfelt thanks, to my wife, daughter and son, your abiding love, active support and constant understanding made this last year bearable. I do not know how to express my gratitude for the countless unspoken sacrifices you made over the last year and a half. My greatest regret at AFIT has been the time I missed with you. Thank you all!

Frank

Table of Contents

List of Figures

List of Tables

Page

AN INVESTIGATION OF THE EFFECTS OF CORRELATION AND AUTOCORRELATION ON CLASSIFIER FUSION WITH NON-DECLARATIONS

1. Introduction

1.1 Background

Correct discrimination of hostile and friend forces is paramount to success in air operations. The possibility of incorrectly classifying hostiles as friends and vice versa can cost the lives of U.S. servicemen. Automatic target recognition (ATR) is one method of discriminating targets. ATR consists of the following six steps: detection, location, combat identification (CID), decision, execution, and assessment (AFPAM 14-210, 1998). Of the six steps necessary for ATR, CID is considered one of the most critical and challenging problems facing the defense community today (Robinson and Aboutalib, 1989). Combat identification accounts for a major point of emphasis in the military's kill chain consisting of search, detect, track, classify, etc. (Haspert, 2000). The use of multi-sensor fusion offers an avenue for improvements in classification accuracy. Multi-sensor fusion combines information from multiple sources to create inferences that cannot be achieved through single source intelligence or information (Hall and Steinberg, 2001). This improvement is supported by Air Force guidance and other research efforts. In particular, Air Force targeting guidance states that the use of fusion should be used whenever possible to enhance intelligence support while adding credibility and accuracy (AFPAM 14-210, 1998). One common assumption in fusion models is independence both between classifiers and across features. Although this occurs in some cases, there is a limited amount of research as to what happens when information collected from

classifiers fails this assumption (Willett, *et al.*,. 2000). Past research and graduate theses investigated the effects of correlation on different fusion methods. This research furthers the previous efforts of Storm *et al.*,. (2003) and Leap *et al.*,. (2004) to include an "indifference zone." The indifference region will consider removal of exemplars within a window corresponding to a high probability of misclassification. Two fusion models will be utilized to test performance with the addition of the indifference region. These models are extensions to the Identification System Operating Characteristic (ISOC) fusion (Haspert, 2002) and Probabilistic Neural Network (PNN) fusion. The ISOC assumes independence of the classifiers while the PNN does not operate on that assumption.

1.2 Problem Statement

In this research the effects of adding a third class, that is non-declared, are explored. Data created in Matlab were used to structure a relatively easily separable problem for a pilot study. Data created in Leap 2004 will be used to test the addition of non-declarations. Several problems will be visited, each with increasing complexity. The desired output will be a useful rule applied to the specific problem geometries considered with the hopes of insight for use of non-declarations in the field. Although training, test and validation data were generated for this research, the methodology should be applicable to real world scenarios.

1.3 Outline of Thesis

This research consists of the following five chapters: Introduction, Literature Review, Methodology, Findings and Analysis, and Conclusions. A succinct chapter by chapter explanation follows.

Chapter 1: Introduction - Background for this research as well as the problem statement and research objectives are developed.

Chapter 2: Literature Review - Reasons to fuse data, methods for fusion of data and their assumptions, and pertinent literatures are reviewed as well as data generated in Leap (2004) used for testing and validation.

Chapter 3: Methodology - Fusion methods employed are addressed including two heuristics for adding non-declarations based on the cost of misclassification.

Chapter 4: Findings and Analysis - Bulk of research showing what happened when problems described in Chapter 3 were actually tested.

Chapter 5: Conclusion and Recommendations - A brief review of the research results as well as ideas for follow on research are presented.

2. Literature Review

2.1 Introduction

This chapter focuses on relevant literature applicable to this research. This review initially discusses the need for classifier fusion as found in Air Force Doctrine. Next, it explores the fusion methods utilized. Then, it details the two types of multi-classifier fusion methods utilized in this study. Finally, the data sets that were employed are described in detail.

2.2 Air Force Direction

"Since the Wright Brothers first flew at Kitty Hawk, the airplane has continually evolved as an instrument of military and national power. Today, the proper employment of aerospace power is essential for success on and over the modern battlefield" (AFDD 2-1, 2000). In order to achieve success, targets must be correctly identified leading to precision engagement of the enemy. Intelligence for targets should be based upon multiple sources for improved accuracy and reliability (AFPAM 14-210). Air force doctrine clearly states that great care should be taken to minimize civilian casualties while military objectives are correctly identified and attacked (AFPAM 14-210, 1998). Minimizing civilian casualties requires sound target intelligence which enhances military effectiveness by showing that the risks undertaken are militarily worthwhile (AFPAM 14-210, 1998). Intelligence, surveillance, and reconnaissance (ISR) are critical aerospace mission areas related to CID. ISR relies on fused information for accurate intelligence suitable to deny adversary efforts at impeding information collection. Fused information shows the big picture allowing commanders a more lucid depiction of the battlespace

(AFDD 2-5.2, 1999). Combining multi-source data into necessary intelligence useful in decision making is called fusion (AFPAM 14-210, 1998).

2.3 Fusion Methods

Identification System Operating Characteristic ISOC fusion and Probabilistic Neural Networks PNN are the two techniques considered in this effort. The main difference in the above methods can be viewed from a top level as the difference between feature level and decision level fusion. Neural networks operate in a manner similar to feature level fusion (see Figure 2-1 below). Features are extracted and fused in the chosen network before an exemplar is classified into a group based upon a probability of class membership. Decision level fusion first labels exemplars at the individual classifier level. These labels are then fused to create a single fused indication for a target. The ISOC fusion method is one example of decision level fusion. Once exemplars are classified into output labels (hostile, friend, etc.), they are compared using logical rules. One simple rule for combining these output labels, assuming only two classifiers, would be to declare a target as hostile only if both classifiers labeled the target as hostile (Robinson and Aboutalib, 1989). Figure 2-1 addresses the difference between feature and decision level fusion in more detail.

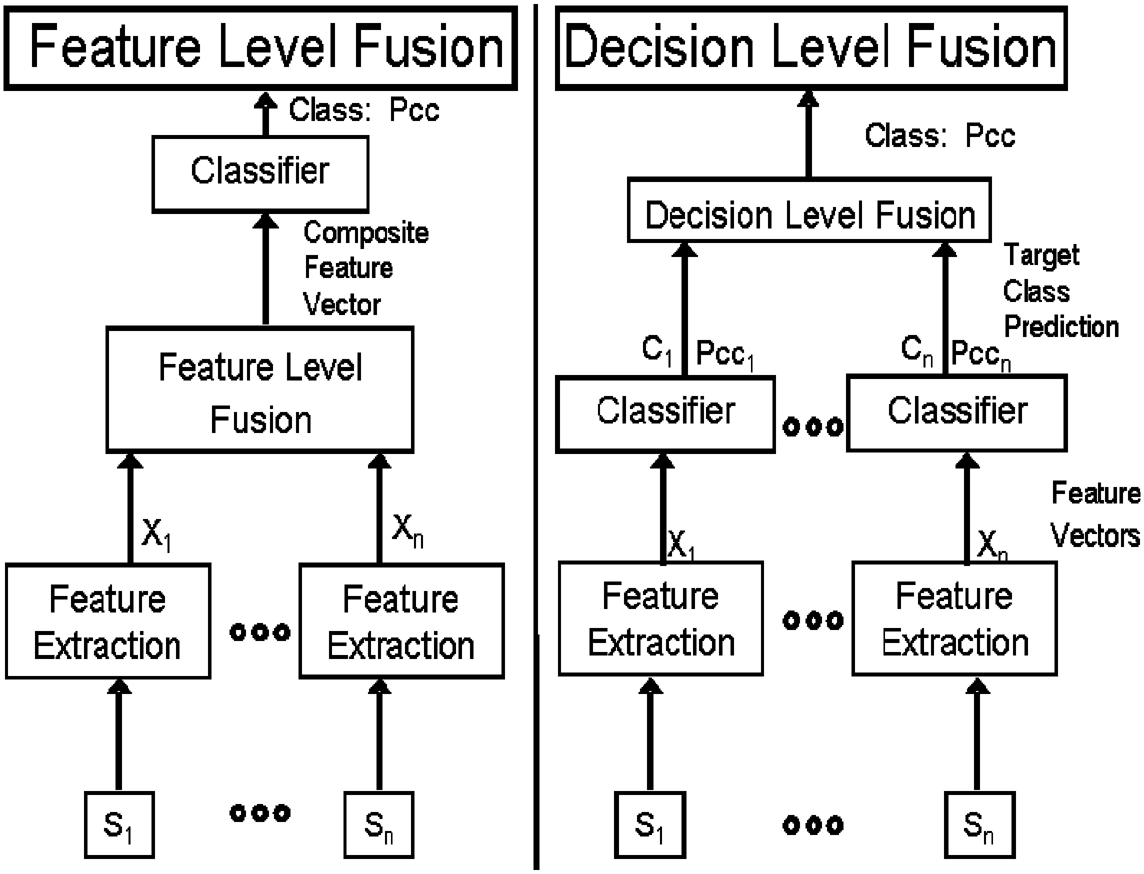


Figure 2-1: Multi-Classifier Fusion Levels (Robinson and Aboutalib, 1989).

2.4 Multi-Classifier Fusion Levels

One major assumption of decision level fusion is independence of classifiers, specifically independence of feature vectors classified by classifiers (Robinson and Aboutalib, 1989). Little is known when the assumption of independence does not hold for a given feature set (Willett, 2000). Robinson and Aboutalib considered the mathematical implications of dependence on decision level fusion techniques. Assume a population set contains two distinct classes, C_1 and C_2. Also assume the *a priori* probabilities of class membership are known, $P(C_1)$ and $P(C_2)$ (Robinson and Aboutalib,

1989). The cost of decisions are as follows: $F(C_1,C_1) = \alpha_{11}$ is the cost of a true positive, $F(C_2,C_2) = \alpha_{22}$ is the cost of a true negative, $F(C_1,C_2) = \alpha_{12}$ is the cost of a false positive, and $F(C_2,C_1) = \alpha_{21}$ is considered the cost of a false negative. Assume that $F(C_i,C_j) > F(C_i,C_i)$ for $i \neq j$ and $i = 1,2$ (Robinson and Aboutalib, 1989). With two features, a composite feature vector can be defined as $\underline{X}_m = \begin{bmatrix} \underline{X}_1 \\ \underline{X}_2 \end{bmatrix}$. Through Baye's rule and the expected value of the cost function, it can be shown that the likelihood of being in C_1 is

$$\frac{p(\underline{X}_m \mid C_1)}{p(\underline{X}_m \mid C_2)} > \frac{(\alpha_{12} - \alpha_{22})}{(\alpha_{21} - \alpha_{11})} \cdot \frac{p(C_2)}{p(C_1)}$$ (Robinson and Aboutalib, 1989).

A decision rule from the above likelihood is thus,

if $\left[\frac{p(\underline{X}_m \mid C_1) \cdot p(C_1)}{p(\underline{X}_m \mid C_2) \cdot p(C_2)}\right] > \frac{(\alpha_{12} - \alpha_{22})}{(\alpha_{21} - \alpha_{11})}$, then X is assigned to C_1.

With regards to feature level fusion, if $\alpha_{11} = \alpha_{22}$ and $\alpha_{12} = \alpha_{21}$, the decision rule above yields the "minimum probability of error decision" (Robinson and Aboutalib, 1989). The right hand side of the inequality becomes a constant allowing a global minimum to be reached.

Continuing with consideration to decision level fusion, the cost function to be minimized now contains two decisions, D_1 and D_2 corresponding to the labels applied at each classifier. The expected value function now carries added terms and grows more complex. In fact, the likelihood decision function is now changed to the following.

$$\text{if } \frac{p(C_1) \cdot p(\underline{X}_1 \mid C_1)}{p(C_2) \cdot p(\underline{X}_1 \mid C_2)} > \frac{\sum_{D_2 = C_1}^{C_2} \gamma_1 \int_{\underline{X}_2} p(D_2 \mid \underline{X}_2) p(\underline{X}_2 \mid \underline{X}_1, C_2) \mathrm{d}\underline{X}_2}{\sum_{D_2 = C_1}^{C_2} \gamma_2 \int_{\underline{X}_2} p(D_2 \mid \underline{X}_2) p(\underline{X}_2 \mid \underline{X}_1, C_2) \mathrm{d}\underline{X}_2} \Rightarrow \underline{X}_1 \text{ from } C_1$$

where

$$\gamma_1 = [F(D_1 = C_1, D_2, A = C_2) - F(D_1 = C_2, D_2, A = C_1)] \text{ and}$$
$$\gamma_2 = [F(D_1 = C_1, D_2, A = C_2) - F(D_1 = C_2, D_2, A = C_2)]$$

(Robinson and Aboutalib, 1989).

By inspection of the above inequality, to reach a global optimum one needs to know $p(D_2|\underline{X}_2)$ and $p(\underline{X}_2|\underline{X}_1,C_2)$ for all $(\underline{X}_2 \mid \underline{X}_1)$ (Robinson and Aboutalib, 1989). Thus, in order to reach a global optimum using decision level fusion with two possible classes, one should not classify the classifiers independently, but should have prior knowledge of one classifier before using the other. However, if the feature vectors are truly independent of one another, the right hand side of the inequality becomes a constant as in feature level fusion and classifying each classifier independently can return the fused global optimum (Robinson and Aboutalib, 1989). The addition of a third unknown class was not addressed and the effects of correlation were uncertain.

Further research by Willett, Swaszek, and Blum considered what level of classifier processing of a Gaussian shift in mean problem was required to reach optimum performance (Willett *et al.*,, 2000). Feature vectors and correlation coefficients were the parameters considered between exemplars. With this problem, difficulties arising from levels of statistical dependence resulted in several complicated rules. After partitioning the space of Gaussian shift in mean problems into three regions named “the good,” “the bad,” and “the ugly,” research reflected optimal rules based upon the partitioned region of use (Willett *et al.*,, 2000). In particular, any problem in “the good” region required statistical independence of feature vectors in order to reach optimality (Willet *et al.*,, 2000). Outside of “the good” region, complex rules and problem specifics define the rule needed which might be able to operate without the assumption of independence (Willett

et al.,, 2000). Feature level fusion does not require independence of their features; there is not much new information received with high levels of correlation and feature selection might be useful.

2.5 ISOC

The Identification System Operating Characteristic (ISOC) fusion method seeks the lowest operational cost for a given threshold (Ralston, 1999). This particular rule is then applied to all future exemplars. The ISOC differs from traditional classifier fusion methods which frequently utilize fixed rules in seeking a minimum cost (Haspert, 2000). While fixed rules remove difficulty in terms of implementation, they often do not reach a global optimum solution (Haspert, 2000). Bayesian techniques have the ability to produce optimal ID classifier fusion rules (Haspert, 2000). Two common target classes are hostile and friend. This research extends to a third target class, unknown.

2.5.1 Classifier Performance Matrices

Classifiers take an exemplar and output a classification label as shown in Table 2-1. Decision level fusion methods such as the ISOC take these matrices from all available classifiers and make a fused decision.

Table 2-1: Single Classifier Confusion Matrix.

<table>
<tr><td></td><td colspan="4">Indication</td></tr>
<tr><td rowspan="3">Truth</td><td></td><td>“H”</td><td>“F”</td><td>“ND”</td></tr>
<tr><td>H</td><td>P(“H”|H)
TP</td><td>P(“F”|H)
FN</td><td>P(“ND”|H)</td></tr>
<tr><td>F</td><td>P(“H”|F)
FP</td><td>P(“F”|F)
TN</td><td>P(“ND”|F)</td></tr>
</table>

The above confusion matrix displays the possible outputs from a single sensor/classifier. The rows in the table correspond to the true states of nature for targets while the columns represent the declaration made by the classifier. For instance, $P(\text{“ND”} \mid H)$ represents the probability of this classifier declaring the target as unknown when it was in fact hostile. This is the confusion matrix used throughout this research with H being hostile, F being friend, and ND being unknown. This matrix can be expanded as needed to match classifier outputs and classes possible. Accurate classifiers will carry large values in $P(\text{“H”} \mid H)$ and $P(\text{“F”} \mid F)$ and small probabilities elsewhere (Haspert, 2000).

2.5.2 Combat Identification System States

Let N_s be the total number of classifiers in a system and i be the classifier in consideration within the system with $1 \le i \le N_s$ (Ralston, 1998). Further, let n_i be the number of classifier states of the i^{th} classifier accounting for the rows of the performance matrix shown in Table 2-2. Finally, let k_i be the classification state of the i^{th} classifier with $1 \le k_i \le n_i$. Assuming there is some level of independence across the classifiers, then $N = \prod_{i=1}^{N_s} n_i$ (Ralston, 1998). The following classifier performance matrix in Table 2-2 represents the probabilities found from data accumulated through exercises, tests or analyses (Ralston, 1998). It can easily be adapted to handle more true classes as well as

output states. Let S_j denote the j^{th} configuration of the combat ID system (CID) where $S = \bigcup_{j=1}^{N} S_j$ represents all possible configurations of the CID with $1 \le j \le N$ (Ralston, 1998).

Table 2-2: Classifier Performance Matrix.

j	S_j
1	$(s_1^1, s_2^1, s_3^1, \ldots, s_{N_s}^1)$
2	$(s_1^2, s_2^2, s_3^2, \ldots, s_{N_s}^2)$
3	$(s_1^3, s_2^3, s_3^3, \ldots, s_{N_s}^3)$
. .	. .
. .	. .
N	$(s_1^N, s_2^N, s_3^N, \ldots, s_{N_s}^N)$

The classifier performance matrix can be used to create the probability matrix from Table 2-1. If there is negligible correlation among classifiers, classifier probability matrices can be multiplied to create conditional probabilities of being in a state of the classifier performance matrix S_j given truth T ($T \in \{H, F\}$). Thus, an equation for this probability is $P(S_j \mid T) = \prod_{i=1}^{N_s} P(s_i^j \mid T)$ where s_j^i denotes the state of the j^{th} classifier in the i^{th} configuration (Leap, 2004; Ralston, 1998). It is important to note that $\sum_j P(S_j \mid T) = 1$

such that $\sum_j P(S_j \mid H) = \sum_j P(S_j \mid F) = 1$ (Ralston, 1998). The classifier performance matrix for a two classifier, three output problem as used in parts of this research would look like the one in Table 2-3.

Table 2-3: Sample Classifier Output State Matrix.

j	(s_1^j, s_2^j)
1	(H,H)
2	(H,U)
3	(H,F)
4	(U,H)
5	(U,U)
6	(U,F)
7	(F,H)
8	(F,U)
9	(F,F)

2.5.3 Identification Fusion Rules

A fusion rule consists of a vector relaying classifier output combinations to declare in a class. This rule is made to resolve conflicting indications from independent classifiers (Ralston, 1998). A complete identification (I.D.) fusion rule can be expressed as a vector R. This N dimensional vector R = (r_1, r_2, r_3, …,r_N) corresponds to the different state outputs as shown in Table 2-3 above. For $j = 1, 2, \ldots, N$, each $r_j \in \{0, 1\}$ denotes a declaration for that rule (Leap, 2004). For instance, in the case considered here, a hostile rule could be $(1, 1, 0, 0)$. This rule would declare a target as hostile “H” any time both classifiers labeled the target as hostile $("H","H")$ or only the first classifier declares hostile $("H","F")$ (Ralston, 1998). The converse will be classified as friends,

$("F","H")$ or $("F","F")$. The probability of declaring a target as hostile given that it is hostile is $P("H"|H) = \sum_{j=1}^{N} P(S_j | H) \cdot r_j$ and the probability of misclassifying the target as a hostile is $P("H"|F) = \sum_{j=1}^{N} P(S_j | F) \cdot r_j$ (Ralston, 1998). Figure 2-2 further depicts the ISOC fusion process.

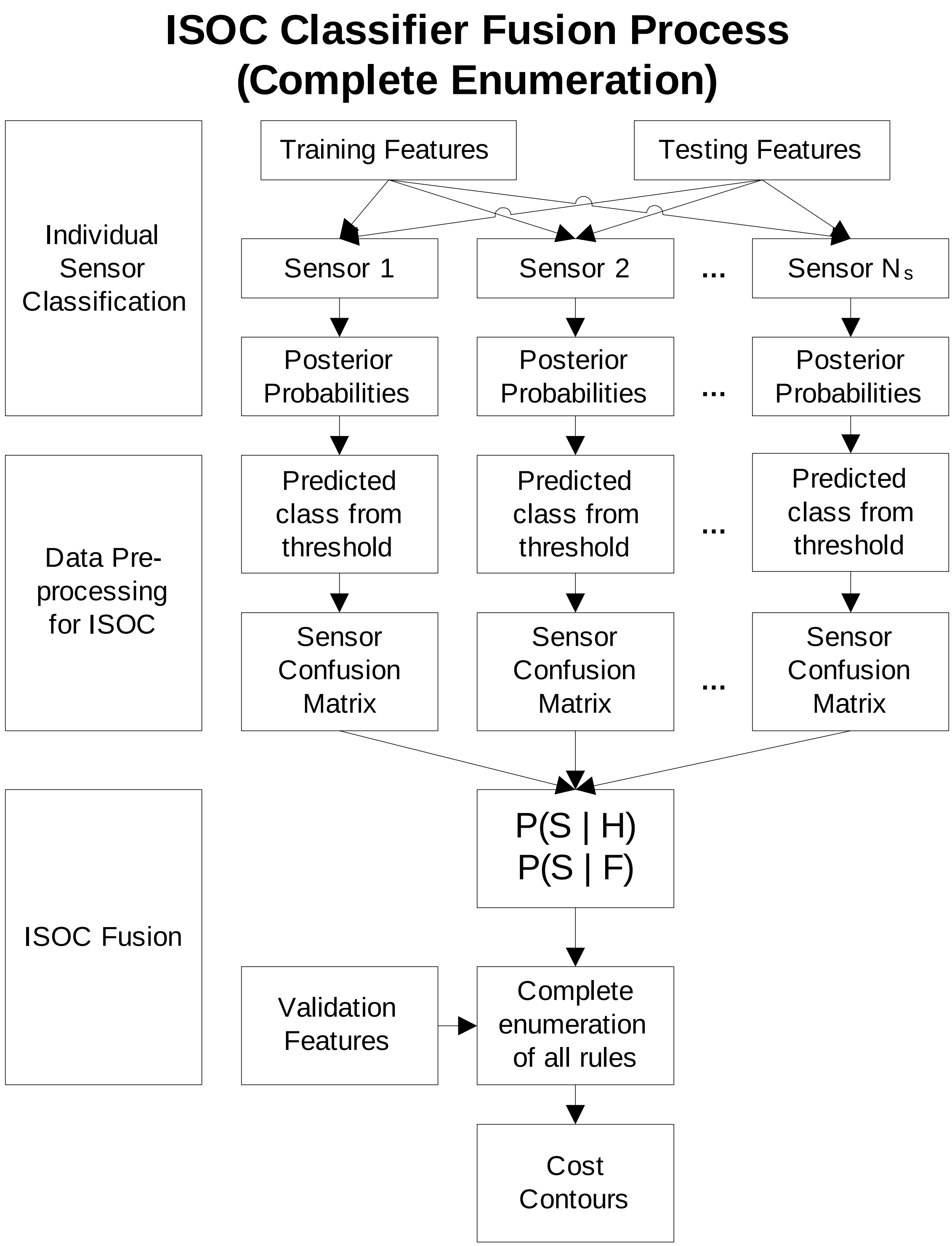


Figure 2-2: ISOC Classifier Fusion Process (Forced Decision)

Complete enumeration of all possible rules can become daunting with the addition of classifiers and/or states. For instance, a two classifier-three output system would contain 512 possible rules (2^9), but a three classifier-three output system would contain 134,217,728 possible rules or (2^{27}). While 2^{27} can require days of run time depending on machine speed, 2^9 rules are manageable and thus complete enumeration was employed for the ISOC fusion forced decision method described in chapter three to find the optimal rule set based on cost.

2.5.4 Likelihood Ratio Approach to Rules Selection

In order to achieve the best classification in the two classifier case, one must look to find the rule that maximizes $P("H"|H)$ while minimizing $P("H"|F)$; a solution where the greatest number of true targets are identified with the least number of false alarms possible for the two class problem would then be reached (Ralston, 1998). The best possible classifier performance would be to correctly identify all hostiles and friends. This is usually not feasible, but the likelihood ratio described below allows a chance to see how close a set of classifiers are to that perfect classification. The likelihood ratio method considers the optimal hostile and friend rules built in sequence element by element. There are two rules that can easily be compared to the optimal classifier performance, never declare any targets hostile and always declare all targets hostile (Haspert, 1998). The "always declare hostile" rule ensures that no targets are missed at the cost of friendly casualties; this rule is accomplished by declaring all exemplars hostile by setting all elements of the hostile rule $r_j = 1$ for all $j = 1, \ldots, N$ (Ralston, 1998). This ensures all hostile forces are engaged at the cost of the highest level of fratricide possible. The most conservative "never declare hostile" rule is found by setting all elements of the

hostile rule $r_j = 0$ for all $\mathrm{j} = 1, \ldots, \mathrm{N}$. The next most conservative rule will contain 1 $r_j = 1$ and all other $r_j = 0$ thus only declaring one classifier output state as hostile. This rule is the state with the largest likelihood ratio found by calculating $\frac{p(S_j \mid H)}{p(S_j \mid F)}$ for all j and sorting these ratios from greatest to smallest. States are added in order of their likelihood ratios to the hosile rule until the "always declare hostile" rule is reached. This process of adding rules based on their likelihood of being hostile forms the optimal ISOC boundary. The following algorithm further explains this ISOC boundary creation (Storm, Bauer and Oxley, 2003; Leap, 2004).

1. Calculate $P(S_j \mid T)$ for all $j = 1,...,9$ and $T \in \{H, F\}$ from the classifier confusion matrices.
2. Calculate $LR^j = P(S_j \mid H) / P(S_j \mid F)$ for all j where LR^j the likelihood ratio for state j of the classifier output matrix is.
3. Rank LR^j from greatest to smallest such that $LR_{[1]}^{j_1} \geq LR_{[2]}^{j_2} \geq \ldots \geq LR_{[N]}^{j_N}$ where $LR_{[1]}^{j_1}$ is the largest and $LR_{[N]}^{j_N}$ is the smallest ratio.
4. Select S_j corresponding to the largest $LR_{[N]}^{j_N}$ not yet included in the hostile fusion rule ($r_{j_N} = 1$ in R)
5. Go to 3 unless $r_j = 1$ for all j

This algorithm "turns on" elements of the hostile rule in decreasing order of their likelihood ratio (Ralston, 1998). The two extreme rules described above generate the end points of the boundary and the above algorithm forms the rest of it in succession of

likelihoods. Thus, there will be $N+1$ distinct rules forming the ISOC boundary (Ralston, 1998). The ISOC fusion process using likelihood ratios is depicted in Figure 2-3.

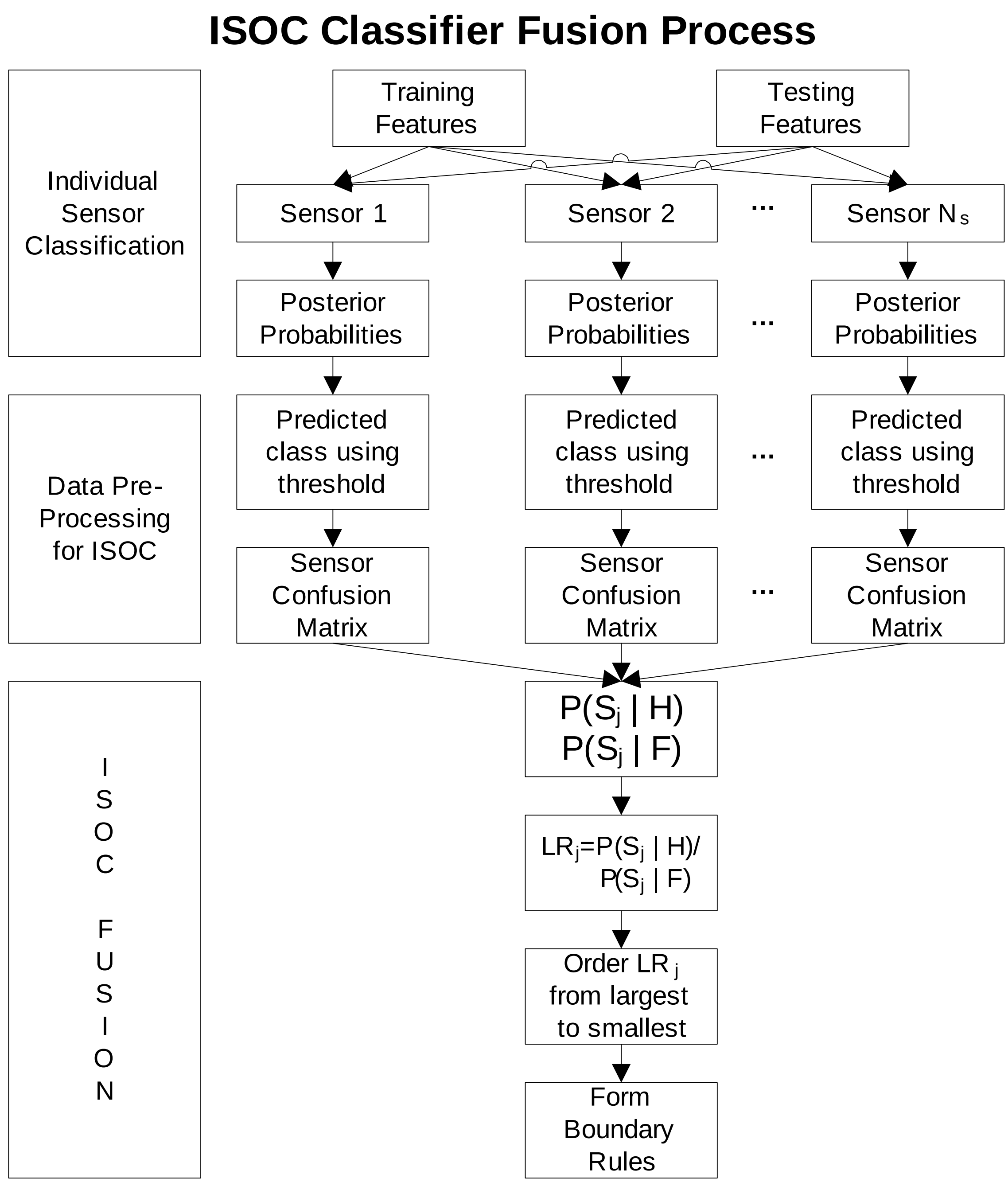


Figure 2-3: ISOC Classifier Fusion by Likelihood Ratios.

This forms the basis for using likelihood ratios to generate optimal rules. This section will be readdressed in chapter three with further implementation concerns.

2.5.5 Total Cost of Misclassification

In order to have a true comparison between rules, a metric must be defined. The metric used herein calculates costs of misclassifications based on probabilities found in the classifier probability matrix in Table 2-1 (total cost function below does not include non-declarations; "ND" in table will be addressed in this context in chapter three). For the two classifier-two output case, a total cost function is

$$C_T = C_{FN} \times P(H) \times P(FN) + C_{FP} \times P(F) \times P(FP) \qquad (2\text{-}1)$$

Where

$C_T =$ Total cost for the rule being tested

$C_{FN} =$ Cost of a false negative

$P(H) =$ a priori probability of a target being hostile

$P(FN) = \sum_{j=1}^{N} P(S_j \mid H) \times \bar{r}_j =$ Probability of a false negative

$r_j =$ Element j of rule R (Defined in 2.5.6)

$\bar{r}_j = 1 - r_j$ (Complement of rule R)

$C_{FP} =$ Cost of a false positive

$P(F) =$ a priori probability of a target being a friend

$P(FP) = \sum_{j=1}^{N} P(S_j \mid F) \times r_j =$ Probability of a false positive

In order to compare rules allowing non-declarations, terms need to be added and will be addressed in the methodology section.

2.5.6 Notation

Thus far, rules were limited to hostile or friend with the following hostile notation:

$R = (r_1, r_2, \ldots, r_N)$ where $r_i \in \{0,1\}$ for $i = 1,2,\ldots,N$ with N representing the possible number of output combinations from classifiers as shown in Table 2-3. It was implicitly assumed that the friend rule was simply the complement to the hostile rule or $\overline{R} = (\overline{r}_1, \overline{r}_2, \ldots, \overline{r}_N)$ where $\overline{r}_i = 1 - r_i$ for $i = 1,2,\ldots,N$. In later sections, this will be extended to a non-declared rule requiring the notation from Ralston (1998) to be expanded as follows. Let H represent the hostile rule where $H = (h_1, h_2, \ldots, h_N)$ with $h_i \in \{0,1\}$; let F denote the friend rule with $F = (f_1, f_2, \ldots, f_N)$ and $f_i \in \{0,1\}$; further, let ND be the non-declared rule with $ND = (nd_1, nd_2, \ldots, nd_N)$ and $nd_i \in \{0,1\}$. These rules are mutually exclusive and collectively exhaustive such that $h_i + f_i + nd_i = 1$ for all $i = 1,2,\ldots,N$. This notation will be further addressed in 3.2.4.

2.6 PNN Fusion Method

A Probabilistic Neural Network (PNN) is a useful tool proven to converge to the Bayesian optimal classifier if given enough data for training (Wasserman, 1993). The PNN trains very quickly and is robust to noise (Wasserman, 1993). The amount of computations required to make classifications with a PNN greatly depend on the size of the training set (Wasserman, 1993). Figure 2-4 shows a probabilistic neural network.

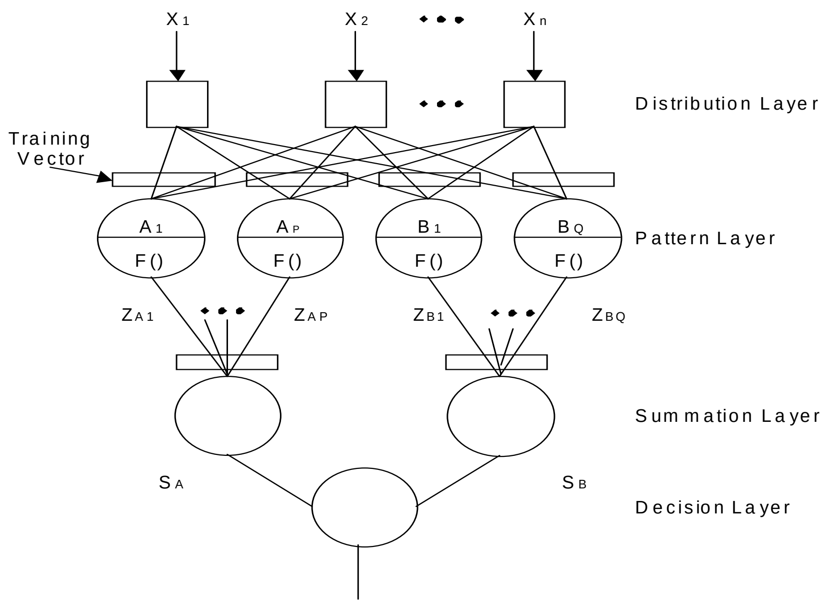


Figure 2-4: Probabilistic Neural Network (Wasserman, 1993)

Figure 2-4 shows a normalized input vector $X = (x1, x2, \ldots, xn)$ fed to the distribution layer of a PNN. The distribution layer is a connection point and does not perform any computations (Wasserman, 1993). The weights applied to each distribution layer vector heading into any one pattern layer correspond to a specific training vector. The pattern layer computes the sum of the weights contributed from every distribution layer neuron and applies to it a non-linear function yielding Z_{ci}; c corresponds to the particular training vector used and i indicates the pattern layer involved in the computation (Wasserman, 1993). Each Z_{ci} is formed by the equation $Z_{ci} = \exp\left[\frac{(X_{Ri}^{t} X_i - 1)}{\sigma^2}\right]$ where X_{Ri} denotes the particular training vector utilized for the pattern layer considered (Wasserman, 1993). The summation layer, related to a particular class, takes all Z_{ci} from its class and computes $S_c = \sum_{i=1} \exp\left[\frac{(X^t X_{Ri} - 1)}{\sigma^2}\right]$ (Wasserman, 1993). For the PNN displayed in

Figure 2-4, there are two possible classes corresponding to the two summation layer neurons. The output from the summation layers are then compared at the decision layer yielding a one if $S_A > S_B$ and a zero if the opposite is true. Designating a target with a one labels it as class A. The link to more classes is simply the addition of pattern layer neurons as well as more summation layers; the determination of class membership is then chosen by the largest summation layer value (Wasserman, 1993).

2.7 Data Generation

The data used for this research were created in Leap 2004. Several different data sets were created to test a broad range of problems. Two discriminant functions were utilized to serve as classifiers, linear and quadratic. The linear classifier was created with assumed equal covariance and equal prior probabilities of class membership. It generated the posterior probabilities of class membership for use in fusion. The equation used for the posterior probabilities in the linear classifier was

$$P(\pi_j \mid X_0) = \frac{e^{[-\frac{1}{2}(\underline{X}_0 - \underline{\mu}_j)' \sum^{-1} (\underline{X}_0 - \underline{\mu}_j)]}}{\sum_{i=1}^{k} e^{[-\frac{1}{2}(\underline{X}_0 - \underline{\mu}_j)' \sum^{-1} (\underline{X}_0 - \underline{\mu}_j)]}} \ .$$

The quadratic classifier operated under the assumptions that the prior probabilities were equal, but the covariance matrices could not be assumed equal. The posterior probability equation used for the quadratic classifier was

$$P(\pi_j \mid X_0) = \frac{\dfrac{1}{(2\pi)^{\frac{p}{2}} \left|\sum_j\right|^{\frac{1}{2}}} e^{[-\frac{1}{2}(\underline{X}_0 - \underline{\mu}_j)' \sum_j^{-1} (\underline{X}_0 - \underline{\mu}_j)]}}{\sum_{i=1}^{k} \dfrac{1}{(2\pi)^{\frac{p}{2}} \left|\sum_i\right|^{\frac{1}{2}}} e^{[-\frac{1}{2}(\underline{X}_0 - \underline{\mu}_i)' \sum_i^{-1} (\underline{X}_0 - \underline{\mu}_i)]}}.$$

Two types of correlation were considered to match possible real world scenarios, inter-correlation and intra-correlation. Inter-correlation describes the correlation between the features of a data set (Leap, 2004). For instance, the height and width of a particular target may have a strong correlation; the length of a T-72 tank is in proportion to its width. Table 2-4 demonstrates inter-correlation.

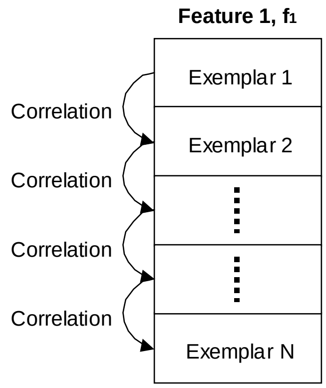
Table 2-4: Correlation table from Leap (2004).

Correlation Correlation Correlation

Feature 1, f_1	**Feature 2, f_2**	**Feature 3, f_3**	**Feature 4, f_4**
Exemplar 1	Exemplar 1	Exemplar 1	Exemplar 1
Exemplar 2	Exemplar 2	Exemplar 2	Exemplar 2
:	:	:	:
Exemplar N	Exemplar N	Exemplar N	Exemplar N

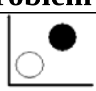
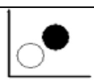
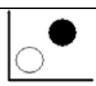
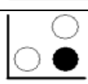
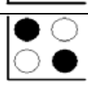
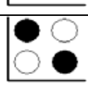
The next type of correlation addressed is intra-correlation or auto-correlation. This explores the relationship between incoming exemplars in a given feature. Table 2-5 depicts the autocorrelation of a feature.

Table 2-5: Correlation table from Leap (2004).



The six problems considered in this research are encapsulated in Table 2-6. The icon shown for each problem depicts the locations of data centroids for the different classes used.

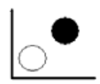
Table 2-6: Problem Description from Leap (2004).

Problem #	Problem Name	Problem Description
1 	4 Feature Case	Recreates Storm work; average cost surface of 5 runs as response
2 	8 Feature Case	Adds noise and redundant features to problem 1; changes mean of class 1
3 	8 Feature with Autocorrelation Case	Adds autocorrelation to problem 2; changes mean of class 1
4 	8 Feature Triangle Case	Changes geometry of problem 2
5 	8 Feature XOR Case	Changes geometry of problem 4
6 	8 Feature XOR with Autocorrelation Case	Adds autocorrelation to problem 5

The above problems are all variants of the same structure of correlation matrices and mean vectors. For all problems considered, $F = F_1 \times F_2 \subset R^n$ where n is the total number of features considered in the problem (assumed to be an even integer). Thus,

$F_1 \subset R^{n/2}$ represents the features observed by classifier 1, the linear discriminant function and $F_2 \subset R^{n/2}$ represents the features observed by classifier 2, the quadratic discriminant function. Problems 5 and 6 created a space that was too difficult for the linear and quadratic classifiers to discriminate. Thus, a probabilistic neural network will also be used as a classifier for these problems and fused with a quadratic function. Assuming only two possible classes, class 0 and class 1, let F_i^j be the features from feature set i in class j where $F_1 = F_1^0 \cup F_1^1$ and $F_2 = F_2^0 \cup F_2^1$ (Leap, 2004). The mean vector for feature set i in class j is represented by μ_i^j. The correlation of the data is given by $\sum^i = \begin{bmatrix} \sum^i_{F_1,F_1} & \sum^i_{F_1,F_2} \\ \sum^i_{F_2,F_1} & \sum^i_{F_2,F_2} \end{bmatrix}$ for all class i (Leap, 2004). For the purposes of all data sets considered, the covariances of the two classes were set equal to each other ($\sum^0 = \sum^1$). The correlation between and within feature sets was created through the use of different ρ values within the correlation matrices. Four different ρ will be addressed: $\rho, \rho_{\text{red}}, \rho_{\text{ind}}$, and ρ_{auto}. The primary correlation is included through ρ. This correlation affects the correlations between features and $\rho \in \{0.0, 0.2, 0.4, 0.6, 0.8, 0.9\}$ (Leap, 2004). The correlation due to the addition of a redundant feature is attributed to $\rho_{red} = 0.5$; correlation induced by the addition of ρ and ρ_{red} is characterized by $\rho_{ind} = \rho \times \rho_{red}$; the correlation within a feature set is described by $\rho_{auto} \in \{0.0, 0.5, 0.9\}$ (Leap, 2004). Generating multivariate normal data with autocorrelation required the use of some equations found in Laine, 2003. Let z(t) where $t \in \{1, 2, ..., N\}$ be an exemplar of the current feature space with N being the total number of exemplars present (Laine, 2003;

Leap, 2004). The distribution of each exemplar is then $z(t) \sim N(\bar{0}, \sum{}^{0})$ (Laine, 2003; Leap, 2004). Letting $A = \rho_{auto} * I, B = (\sqrt{1-\rho_{auto}^{2}}) * I$, and $\varepsilon(\mathrm{t}) \sim \mathrm{N}(\bar{0}, (B * \sum{}^{0} * B))$ for each exemplar t allows the following to hold: $z(t) = A * z(t-1) + \varepsilon(t)$ (Laine, 2003; Leap, 2004). Problems 5 and 6 were too difficult for a linear discriminant function to separate and they were revisited using a PNN and quadratic discriminant function as classifiers.

2.7.1 Problem 1: 4 Feature Case 

This is the only problem considering 4 features. Let $F_1 \subset R^2$ be the features observed by classifier 1 and $F_2 \subset R^2$ be the features observed by classifier 2. It follows that $F = F_1 \times F_2 \subset R^4$. The data was created such that the features within individual feature sets were statistically independent as shown by $\sum_{F_1,F_1}^{i} = \sum_{F_2,F_2}^{i} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ (Leap, 2004). The correlation matrix between features in F_i,F_j where $i \neq j$ is represented by $\sum_{F_1,F_2}^{i} = \sum_{F_2,F_1}^{i} = \begin{bmatrix} 0 & \rho \\ \rho & 0 \end{bmatrix}$ where $\rho \in \{0.0, 0.2, 0.4, 0.6, 0.8, 0.9\}$ (Leap 2004). The mean vectors for all classes and feature sets are as follows:

$\mu_1^0 = (0,0)^T$, $\mu_1^1 = (0.95, 0.95)^T$, $\mu_2^0 = (0,0)^T$, $\mu_2^1 = (1.15, 1.15)^T$. The feature sets for problem 1 are distributed as follows:

$$F_1^0 \sim N(\mu_1^0, \sum_{F_1,F_1}^{0}), F_1^1 \sim N(\mu_1^1, \sum_{F_1,F_1}^{1}), F_2^0 \sim N(\mu_2^0, \sum_{F_2,F_2}^{0}), F_2^1 \sim N(\mu_2^1, \sum_{F_2,F_2}^{1}).$$

2.7.2 Problem 2: 8 Feature Case

This problem adds a feature vector to problem 1 as noise variables and another feature vector as redundant features. The class 1 mean vector is also changed as reflected below:

$\mu_1^0 = (0,0,0,0)^T$, $\mu_1^1 = (0.5,0.5,0.5,0)^T$, $\mu_2^0 = (0,0,0,0)^T$, $\mu_2^1 = (0.75,0.75,0.75,0)^T$. The feature vector distributions remain unchanged symbolically with the only changes occurring due to the altered class 1 mean vector and the additional features. The correlation due to $\sum_{F_1,F_1}^i = \sum_{F_2,F_2}^i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & \rho_{red} & 0 \\ 0 & \rho_{red} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ and

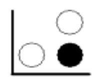
$\sum_{F_1,F_2}^i = \sum_{F_2,F_1}^i = \begin{bmatrix} 0 & \rho & \rho_{ind} & 0 \\ \rho & 0 & 0 & 0 \\ \rho_{ind} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ comprises the overall correlation (Leap, 2004).

The $\rho_{ind} = \rho * \rho_{red}$ comes about due to the fact that feature 2 and feature 3 are correlated by design as well as feature 6 and feature 7 as shown by $\sum_{F_1,F_1}^i$ and $\sum_{F_2,F_2}^i$. The addition of ρ between features 1 and 6 causes an induced correlation between features 1 and 7 with a similar occurrence between features 3 and 5 due to correlation between features 2 and 5.

2.7.3 Problem 3: 8 Features with Autocorrelation Case

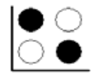
This problem adds autocorrelation to problem 2. The mean vectors were also varied in the hopes of covering a broader spectrum of problem types. For this problem,

the mean vectors are $\mu_1^0 = (0,0,0,0)^T$, $\mu_1^1 = (0.95,0.95,0.95,0)^T$, $\mu_2^0 = (0,0,0,0)^T$, and $\mu_2^1 = (1.15,1.15,1.15,0)^T$ respectively. The covariance matrices remain unchanged from problem 2. The distributions of the feature vectors are the same symbolically with the only change present due to the new mean vectors. The levels of correlation are the same as problem 2 such that $\rho \in \{0.0,0.2,0.4,0.6,0.8,0.9\}$, $\rho_{red} = 0.95$, and $\rho_{ind} = \rho * \rho_{red}$ (Leap, 2004). Autocorrelation is made present within each feature set at $\rho_{auto} \in \{0.0,0.5,0.9\}$ covering a low, medium, and high setting of within correlation.

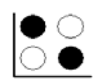
2.7.4 Problem 4: 8 Feature Triangle Case 

The triangle problem varies the geometry of the problem to consider yet another extended environment. There are four multivariate normal populations created, two from class 0 and two from class 1, with three different mean vectors (Leap, 2004). The problem complexity rises making it a little more difficult to separate with a simple discriminant function. The covariance matrices of the data remain the same based on 2 independent features in each class, 1 redundant feature and 1 noise feature independent of all other features. The feature vectors of these new data sets are defined to be F_i^{jk} where i is the feature set number, j is the class, and k is the geometric location of the group of features (i = 1,2, j = 0,1, k = 1,2) (Leap, 2004). For instance, F_1^{01} corresponds to the first set of features from feature set 1 in class 0 and F_1^{02} corresponds to the second set of features from feature set 1 in class 0. Thus the feature vectors now become $F_i^j = F_i^{j1} \cup F_i^{j2}$ and $F_i = F_i^0 \cup F_i^1$ (Leap, 2004). The feature vectors are now distributed as $F_i^{jk} \sim N(\mu_i^{jk}, \sum_{F_i,F_i}^{j})$ for all i, j, k. The mean vectors reflect the same

symbolic notation and are changed to be $\mu_1^{01} = (0,0,0,0)^T$, $\mu_1^{02} = (0.95,0.95,0.95,0)^T$, $\mu_1^{11} = (0.95,0,0,0)^T$, $\mu_1^{12} = (0.95,0,0,0)^T$ for feature set 1 and $\mu_2^{01} = (0,0,0,0)^T$, $\mu_2^{02} = (1.15,1.15,1.15,0)^T$, $\mu_2^{11} = (1.15,0,0,0)^T$, $\mu_2^{12} = (1.15,0,0,0)^T$ for feature set 2 (Leap, 2004).

2.7.5 Problem 5: 8 Feature XOR Case 

This again alters the geometry of problem 3. Problem 3 also generated four populations, but now all four have different mean vectors making a linear discriminant function alone useless. As shown in the icon, four multivariate normal populations will be generated with equal covariance matrices and different means. All of the specifics to this problem are the same as the triangle problem except for the change in mean vectors. The mean vectors have now been changed to the following: $\mu_1^{01} = (0,0,0,0)^T$, $\mu_1^{02} = (0.95,0.95,0.95,0)^T$, $\mu_1^{11} = (0,0.95,0.95,0)^T$, $\mu_1^{12} = (0.95,0,0,0)^T$, $\mu_2^{01} = (0,0,0,0)^T$, $\mu_2^{02} = (1.15,1.15,1.15,0)^T$, $\mu_2^{11} = (0,1.15,1.15,0)^T$, and $\mu_2^{12} = (1.15,0,0,0)^T$ (Leap, 2004).

2.7.6 Problem 6: 8 Feature XOR with Autocorrelation Case 

This problem aims at more extended environments through the use of autocorrelation added to an already difficult discrimination problem. The XOR case from problem 5 above is now altered to consider correlation within a feature set along with correlation across features as visited previously. There are four multivariate normal populations with the same covariance matrices and mean vectors as in problem 5. The only change is the addition of autocorrelation. The addition of autocorrelation is

accomplished in the same manner as problem 3. The within correlation is once again set to $\rho_{auto} \in \{0.0, 0.5, 0.9\}$.

3. Methodology

3.1 Introduction

This chapter details how the research was conducted. First, the experimental design will be described through the use of an example problem. Next, a new cost function will be developed for total cost while considering non-declarations. Then, a heuristic for creating a non-declaration rule from the forced decision ISOC fusion method will be demonstrated. This forced decision heuristic will be continued to the ISOC non-declaration political correctness method. The boundary rules for ISOC fusion will be addressed through a likelihood ratio heuristic. Finally, the PNN fusion method will be revisited to include the new total cost function.

3.2 Experimental Design

This research effort attempted to study the effects of correlation and autocorrelation on classifier fusion when non-declarations were introduced. Two true states of nature were considered, hostile and friend, with equal *a priori* probabilities of each state. Non-declarations were introduced both at the individual classifier level as well as at the fused classification level. At the classifier level, a posterior probability threshold of $T = 0.5$ was set and an indifference window δ_j^i was introduced to allow classifiers to non-declare exemplars that had a high probability of misclassification; the indifference window was denoted $(T \pm \delta_j^i)$ where $\delta_j^i = (i-1) \times 0.05$ for $j = 1,2$ and $i = 1,2,\ldots,11$; δ_j^i denotes the size of the window for classifier j in the i^{th} configuration. Thus, exemplars were classified in the following manner. Let P_k be the posterior probability associated with exemplar k being a true hostile where $k = 1,2,\ldots \# exemplars$; further, let "H" denote the times when an exemplar is labeled hostile, "F" indicate a friend classification and "ND"

denote undeclared. When $i = 1, \delta_j^1 = 0$ and the classifier j labels for exemplar k were forced to “H” or “F” based on the threshold T and any exemplar equal to T was declared “F”; when $i = 11, \delta_j^{11} = 0.5$ and all classifier j labels were forced to “ND” with no hostile or friend declarations; otherwise, labels were determined based upon the following:

$$\text{if} \begin{cases} P_k > T + \delta_j^i, \text{then the exemplar is labeled "}H\text{"} \\ P_k < T - \delta_j^i, \text{then the exemplar is labeled "}F\text{"} \\ \text{else the exemplar is labeled "ND"} \end{cases}.$$

Figure 3-1 shows an individual classifier indifference window. The posterior probability of being in a given class resembles normal distributions with equal variances, but this method of classification could be performed on most distributions.

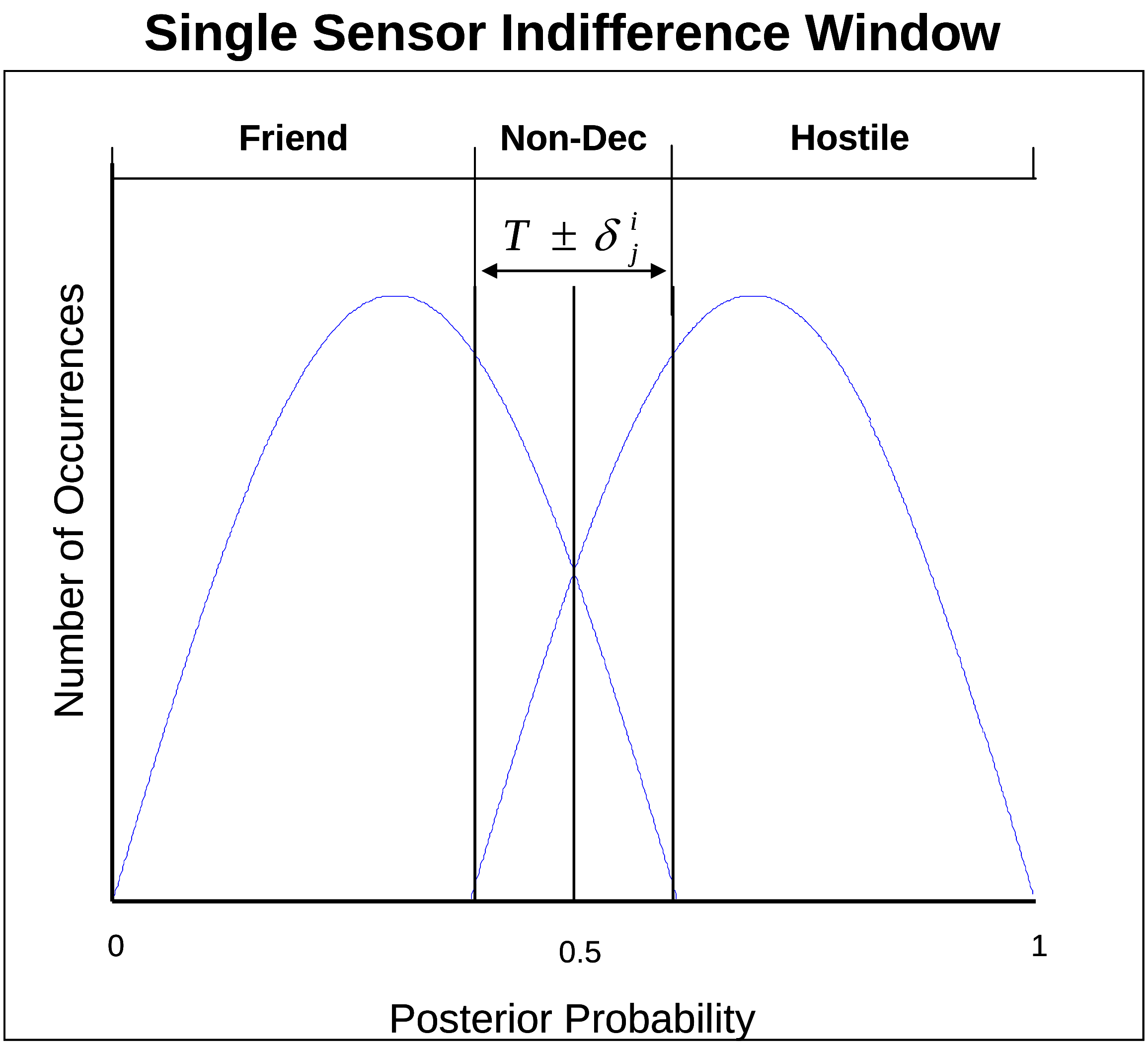
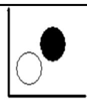
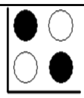
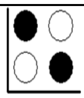


Figure 3-1: Individual Classifier Indifference Window.

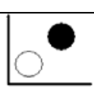
Non-declarations were also allowed from the fused classifier output through the heuristics described later in this section. For the PNN, an indifference window was introduced after the features had been fused. Table 3-1 summarizes the design considerations used in ISOC forced decision (IFD), ISOC non-declaration political correctness heuristic (IND_{PC}) and PNN fusion with non-declarations NF_{ND}.

Table 3-1: Design Considerations

Problem #	Considerations	# Data Sets	Classifiers Used
1 	Sample Size, inter-correlation	15; 3 generated sets for 5 random seeds	Linear/Quadratic
2 	Sample Size, inter-correlation	15; 3 generated sets for 5 random seeds	Linear/Quadratic
3 	Sample Size, inter-correlation, autocorrelation	15; 3 generated sets for 5 random seeds	Linear/Quadratic
4 	Sample Size, inter-correlation	15; 3 generated sets for 5 random seeds	Linear/Quadratic
5 	Sample Size, inter-correlation	15; 3 generated sets for 5 random seeds	Linear/Quadratic PNN/Quadratic
6 	Sample Size, inter-correlation, autocorrelation	15; 3 generated sets for 5 random seeds	Linear/Quadratic PNN/Quadratic

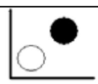
The ISOC non-declaration likelihood ratio heuristic (IND_{LR}) described later held sample size, problem, and correlation levels constant. The heuristic was meant to find the optimal rules set by ratios rather than complete enumeration. Table 3-2 summarizes the considerations for the second heuristic.

Table 3-2: ISOC Non-Declaration Heuristic (IND_{LR}) Considerations

Problem #	Considerations	# Data Sets	Classifiers Used
3 	Sample size = 250, $\rho = \rho_{auto} = 0$	15; 3 generated sets for 5 random seeds	Linear/Quadratic

The final experiment conducted was a full factorial design considering the potential interactions between subjective costs and chosen correlation levels. Table 3-3 summarizes the design considerations.

Table 3-3: RSM Design Considerations

Problem #	Considerations	# Data Sets	Classifiers Used
3 	Sample size = 250, $\rho = 0{,}0.8$, $\rho_{auto} = 0{,}0.9$, $C_{FP} = 10{,}20$, $C_{FN} = 5{,}9$, $C_{ND} = 1{,}4$	15; 3 generated sets for 5 random seeds	Linear/Quadratic

3.2.1 Test Problem

An easily separable problem was desired to test the methodology before extending to the more difficult problems analyzed in chapter 4. This problem will be referenced throughout this chapter to further explain methods and logic. The test problem was created to mirror the experimental design laid out above. To ensure small sample size problems were not encountered, a sample size of 100 exemplars per class was chosen. The problem was designed such that

$$\mu_1 = [3 \quad 3], \sum{}_1 = \begin{bmatrix} 1 & 1.5 \\ 1.5 & 3 \end{bmatrix}, \mu_2 = [6 \quad 2.75], \sum{}_2 = \begin{bmatrix} 1 & -1.5 \\ -1.5 & 3 \end{bmatrix}.$$

This geometry created a problem that allowed both a linear and quadratic discriminant function to easily separate classes to a reasonable accuracy.

3.2.2 ISOC Forced Decision Fusion (IFD)

The first step in considering non-declarations was to allow individual classifiers to label exemplars as unknown. In order to model non-declarations, a grid of possible indifference windows was created for each classifier with the threshold set at T = 0.5. The indifference window was then $T \pm \delta_j^i$ where δ_j^i denotes the size of the window for classifier j in the ith configuration and $\delta_j^i = (i-1) \times 0.05$ for $j = 1,2$ and $i = 1,\ldots,11$. Individual classifier labels were applied to posteriors as described in section 3.2 above. Once all exemplars were labeled for each classifier, the labels were placed in a combined classifier performance matrix in preparation for complete enumeration using ISOC fusion. Next, all possible combinations of hostile and friend rules were tested and fused indications were forced to be "H" or "F". The total cost equation (2-1) had to be modified to consider the probability of non-declarations and their associated cost. This added cost would be a constant for a given grid point of (δ_1^i, δ_2^i) settings and would only affect the overall minimum cost when compared over the range of (δ_1^i, δ_2^i). The added cost due to non-declarations remained a constant for a specified grid point because non-declarations were only allowed at the individual classifier level; the posterior probability of a classifier classifying a target as non-declared was used in the calculation for total cost under the ISOC forced decision method. Under the assumption of independence of the feature sets sent to each classifier, the probability of non-declaration can be calculated through the union of the events that the linear function classifies the target as unknown or the probability that the quadratic function classifies the target as unknown. Thus, it can be shown that

$$P(ND) = P(ND_1) + P(ND_2) - P(ND_1) * P(ND_2)$$

where

P(ND) = Probability that a target is labeled unknown by either classifier

ND_i = a target labeled non-declared from classifier i

$P(ND_i) = P(ND_i \mid H) \times P(H) + P(ND_i \mid F) \times P(F)$ for classifier i.

Note that $P(ND_1)$ and $P(ND_2)$ are simply the probabilities that the specified classifier non-declares an exemplar for a given grid point (δ_1^i, δ_2^i). There are no non-declarations allowed from the overall fusion process at this point, only at the individual classifier classification level. The new total cost function in equation (3-1) for IFD is then constant for each grid point location.

$$C_{T,F} = C_{FN} \times P(H) \times P(FN) + C_{FP} \times P(F) \times P(FP) + C_{ND} \times P(ND) \qquad (3\text{-}1)$$

Now that a cost function is created, total cost must be calculated across all grid points for all possible hostile rules $\delta_1^i \times \delta_2^i \times 512$ with δ_j^i denoting the size of the indifference window on classifier j and 512 representing the total number of hostile rules possible. Since fused non-declarations were not allowed from the IFD, the associated friend rule to a given hostile rule is simply the complement. This ISOC forced decision process yields a set of optimal hostile and friend rules based on minimizing cost for each grid point. IFD was conducted on all of the problems shown in Table 3-1.

3.2.3 ISOC Non-Declaration Heuristic for Political Correctness (IND_{PC})

The above ISOC forced decision method was extended to allow a fused non-declared indication. Once non-declarations were allowed as an output from the fusion process, the total cost function from equation (3-1) was modified to be

$$C_{T,\,ND} = P(FN) \times C_{FN} \times P(H) + P(FP) \times C_{FP} \times P(F) + C_{ND} \times [P(ND_H) \times P(H) + P(ND_F) \times P(F)] \qquad (3\text{-}2)$$

where $C_{T,\,ND}$ = Total cost for the rule being tested allowing non-declarations

C_{FN} = 5, Cost of a false negative

$P(H)$ = 0.5, *a priori* probability of a target being hostile

$P(FN) = \sum_{j=1}^{N} P(S_j \mid H) \times f_j$ =Probability of a false negative

C_{FP} = 10, Cost of a false positive

$P(F)$ = 0.5, *a priori* probability of a target being a friend

$P(FP) = \sum_{j=1}^{N} P(S_j \mid F) \times h_j$ = Probability of a false positive

C_{ND} = 1, Cost of a non-declaration

$P(ND_T) = \sum_{j=1}^{N} P(S_j \mid T) \times nd_j$ = probability of non-declaring classifier j given truth

$T \in \{H, F\}$

The above costs were set by a subject matter expert; they will be used throughout this research unless otherwise stated. The heuristic in Figure 3-2 was developed to incorporate the new total cost function and find an optimal rule set by filtering the optimal rules realized through ISOC fusion with a forced decision.

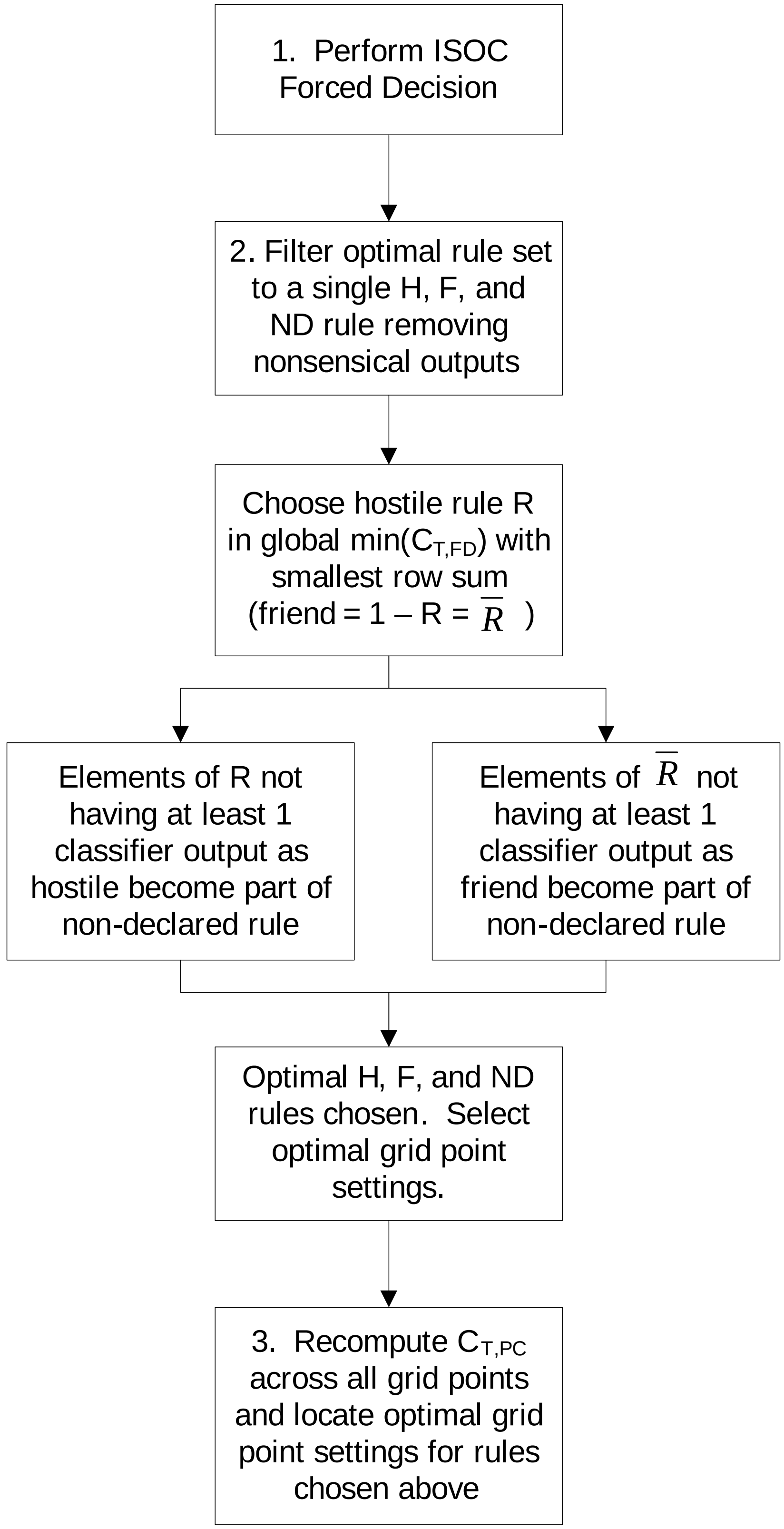


Figure 3-2: ISOC Non-Declaration Heuristic IND_{PC}.

The first step in the heuristic is to perform complete enumeration of ISOC forced decision fusion as explained in 3.2.2 above. This results in an optimal minimum cost for each grid point and all possible rules combinations that can reach that minimum cost. Table 3-4 shows the cost results for step 1 of the example problem.

Table 3-4: Example Problem Grid Point Min Costs.

δ_1	δ_2	$\min(C_{T,FD})$
0	0	0.575
0	0.05	0.59
.	.	.
.	.	.
.	.	.
0.05	0.1	0.4341
0.05	0.15	0.4085
0.05	0.2	0.4231
0.05	0.25	0.4878
.	.	.
.	.	.
.	.	.
0.5	0.45	1.425
0.5	0.5	2.5

Once the global minimum cost is located, step 2 can be performed. This global minimum generally yields several possible hostile and friend rule combinations for the specified grid point. These alternate optimals are found because different combinations of declarations can yield the same cost. These rules are then filtered to a single hostile, friend and non-declared rule. To accomplish this, calculate a row sum for all of the possible hostile rules in the specified grid point. Choose the hostile rule with the smallest row sum and the associated friend rule (in the case of a tie, the first rule with the smallest row sum is selected). The elements of these are then checked to ensure they are practical.

Any hostile rule elements that do not receive at least one “H” output from a classifier become part of the non-declared rule; the friend rule is filtered into a friend and non-declaration rule in the same manner. It would not be very reassuring to have a fused “H” declaration resulting from two classifiers both classifying a target as “F”. You could call this a *political correctness* heuristic where the elements which are counterintuitive become part of the non-declared rule. Figure 3-3 illustrates this step in the heuristic for the example problem.

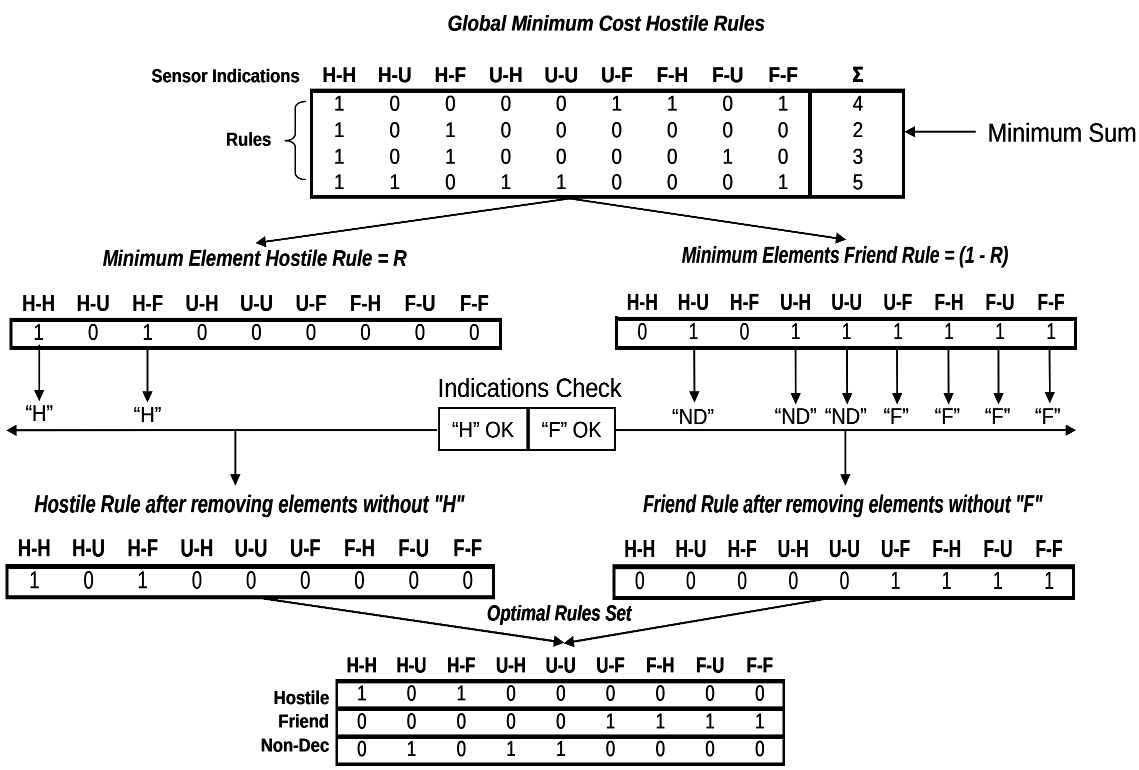


Figure 3-3: Example Problem Rules Selection.

Now that the preferred rules are chosen, $C_{T,PC}$ must be tested at all grid points to locate the optimal settings based upon the rules chosen. Thus, using the new cost function

$C_{T,PC}$ above, the entire grid space is retested and new costs are calculated. The costs achieved through the heuristic are always at least equal to and most often improved from the forced declaration as shown in Table 3-5.

Table 3-5: Example Problem Cost Recalculation.

δ_1	δ_2	$\min(C_{T,FD})$	$\min(C_{T,PC})$
0	0	0.58	0.58
0	0.05	0.59	0.58
.	.	.	.
.	.	.	.
.	.	.	.
0.4	0.2	0.81	0.26
0.4	0.25	0.77	0.25
0.4	0.3	0.78	0.25
0.4	0.35	0.78	0.26
0.4	0.4	0.75	0.25
0.4	0.45	0.79	0.26
0.4	0.5	1.63	0.71
.	.	.	.
.	.	.	.
.	.	.	.
0.5	0.45	1.43	0.63
0.5	0.5	2.50	1.00

The maximum cost shown in Table 3-5 for ISOC forced decision at a given grid point is 2.5 which occurs when all exemplars are classified as friends. The maximum cost for ISOC non-declaration political correctness heuristic (IND_{PC}) is 1 occurring when all exemplars are given non-declared labels. These maximum costs represent the largest minimum cost observed for a given grid point. The minimum cost is bounded for a grid point by these maximum costs.

3.2.4 Likelihood Ratio ISOC Non-declaration heuristic ($\mathbf{IND_{LR}}$)

The heuristic that follows is an extension to the likelihood ratio method described in 2.5.4. This method attempts to find the optimal set of hostile, friend and non-declared rules for a given grid point (δ_1, δ_2). The first step is to calculate the likelihood of being a true hostile by dividing the probability of being a true hostile over the probability of being a true friend $P(S_j \mid H)/P(S_j \mid F) = LR_j$. One difficulty with this step is that some elements j of either $P(S_j \mid H)$ or $P(S_j \mid F)$ and sometimes both might equal zero creating either a divide by 0 situation or an indeterminate form $\left(\frac{0}{0}\right)$. This can be fixed by making the following assumptions. The first is that any time $P(S_j \mid H) = P(S_j \mid F) = 0$, this combination of outputs from the classifiers has not occurred and there is no information as to which declaration would make the most sense. In these cases, it is assumed that these states become part of the non-declared rule ND. This will not affect the value of the cost function while ensuring logical declarations are made. The next two occurrences need to be addressed a little differently. First, consider $P(S_j \mid H) = 0$ and $P(S_j \mid F) > 0$. In this case, the ratio can be computed and the $LR_j = 0$ for all such occurrences. This makes state j highly likely of being a friend which is reasonable since there is no estimated probability that this occurrence will be a hostile. But, when $P(S_j \mid H) > 0$ and $P(S_j \mid F) = 0$, the ratio cannot be calculated. It seems the most reasonable label for this state would hostile. This was accomplished in this research by making a temporary matrix of $P(S_j \mid F)$ and setting all states with zero probability of being a friend equal to ε, the smallest possible number in Matlab. This caused the ratio to be extremely large

making this state the most likely to be hostile. For instance, Table 3-6 shows a notional example where all three of the special cases detailed above occurred.

Table 3-6: Classifier Performance Matrix Example.

State j	Classifier 1 Indication	Classifier 2 Indication	P(S|H)	P(S|F)
1	H	H	0.46	0.02
2	H	U	0.15	0.06
3	H	F	0	0.12
4	U	H	0.23	0
5	U	U	0	0
6	U	F	0.06	0.09
7	F	H	0.06	0.07
8	F	U	0.03	0.15
9	F	F	0.01	0.49

Let $LR_j = LR_{j,[i]}$ represent the likelihood ratio for state j with [i] denoting the i^{th} order statistic for LR_j. The output state j = 4 (U-H) where classifier 1 declared the target as unknown and classifier 2 declared it hostile only occurred on exemplars that were true hostiles. The likelihood ratio for this state was calculated by using ε such that $LR_4 = LR_{4,[1]}$ became the largest ordered likelihood ratio. On the other hand, for output state j = 3 (H-F) when classifier 1 declared the exemplar hostile and classifier 2 declared the exemplar friend, there were no occurrences of this state when an exemplar was a true hostile. The corresponding likelihood ratio became $LR_3 = LR_{3,[8]}$ tying state 6 for the smallest likelihood ratio of being hostile. There were no instances in which state 5 (U-U) occurred so its likelihood ratio was $LR_5 = LR_{5,[9]}$ and it will become part of the non-declared rule. Once the states with no recorded instances of being friend were changed

to ε , the likelihood ratio could be calculated and sorted from largest to smallest. Table 3-7 shows the sorted likelihood ratios from the notional example above. Note that the temporary $\sum P(S_j \mid F) \neq 1$, but this is only used as a proxy to calculate the likelihood ratios.

Table 3-7: Sorted Likelihood Ratios Example.

State j	Classifier 1 Indication	Classifier 2 Indication	P(S|H)	P(S|F)	temp P(S|F)	LR_j
4	U	H	0.23	0	ε = 1E-210	2.3E+209
1	H	H	0.46	0.02	0.02	23
2	H	U	0.15	0.06	0.06	2.5
7	F	H	0.06	0.07	0.07	0.857143
6	U	F	0.06	0.09	0.09	0.666667
8	F	U	0.03	0.15	0.15	0.2
9	F	F	0.01	0.49	0.49	0.020408
3	H	F	0	0.12	0.12	0
5	U	U	0	0	ε = 1E-210	0

Now that the states have been sorted according to their likelihood ratios (hostile to friend), the total cost of misclassification can be calculated; the likelihood ratio of being a friend is calculated by $P(S_j \mid F)/P(S_j \mid H)$. The total cost function from equation (3-2) applies to this heuristic. For the example given in this section, the costs and prior probabilities were assumed to be the following: C_{FN} = 5, C_{FP} = 10, $P(H) = P(F) = 0.5$. Let $P_{cum}(S_{[j]} \mid H)$ represent the cumulative probability of a state given hostile based on the likelihood ratio order statistics above. The probabilities for false positives and false negatives were calculated as follows: $P(FN_{[j]}) = (1 - P_{cum}(S_{[j]} \mid H))$ and $P(FP_{[j]}) = P_{cum}(S_{[j]} \mid F)$ where j represents state j of the classifier performance matrix and

$P_{cum}(S_{[j]}|T)$ is the cumulative probability of state [j] given truth $T \in \{H, F\}$. The costs associated with the sorted classifier performance matrix are shown in Table 3-8.

Table 3-8: Sorted Classifier Performance Matrix Costs.

State j	Classifier 1 Indication	Classifier 2 Indication	P(S|H)	P(S|F)	P_{cum}(S|H)	P_{cum}(S|F)	C_T
4	U	H	0.23	0	0.23	0	1.925
1	H	H	0.46	0.02	0.69	0.02	0.875
2	H	U	0.15	0.06	0.84	0.08	0.8
7	F	H	0.06	0.07	0.9	0.15	1
6	U	F	0.06	0.09	0.96	0.24	1.3
8	F	U	0.03	0.15	0.99	0.39	1.975
9	F	F	0.01	0.49	1	0.88	4.4
3	H	F	0	0.12	1	1	5
5	U	U	0	0	1	1	5

The total cost of misclassification (C_T) in Table 3-8 shows the top three states are to be entered in the hostile rule leaving the remaining states for declaration as friend. This is the least expensive set of rules possible while only declaring targets friend or hostile (currently $nd_j = 0$ for all $j = 1, 2, \ldots, N$). The heuristic to this point has followed the logic of Ralston (1998). From the above example, the associated hostile and friend rules are shown in Table 3-9.

Table 3-9: Optimal Hostile and Friend rules.

State j	Classifier 1 Indication	Classifier 2 Indication	P(S|H)	P(S|F)	C_T	Hostile	Friend	H	F
4	U	H	0.23	0.00	1.93	1	0	$h_{[1]}$	$f_{[1]}$
1	H	H	0.46	0.02	0.88	1	0	$h_{[2]}$	$f_{[2]}$
2	H	U	0.15	0.06	0.80	1	0	$h_{[3]}$	$f_{[3]}$
7	F	H	0.06	0.07	1.00	0	1	$h_{[4]}$	$f_{[4]}$
6	U	F	0.06	0.09	1.30	0	1	$h_{[5]}$	$f_{[5]}$
8	F	U	0.03	0.15	1.98	0	1	$h_{[6]}$	$f_{[6]}$
9	F	F	0.01	0.49	4.40	0	1	$h_{[7]}$	$f_{[7]}$
3	H	F	0.00	0.12	5.00	0	1	$h_{[8]}$	$f_{[8]}$
5	U	U	0.00	0.00	5.00	0	1	$h_{[9]}$	$f_{[9]}$

Notice that the state when both classifiers non-declare a target (U-U) falls into the friend rule for now, but it accounts for zero cost. This element will become part of the non-declared rule at the end of this heuristic. The double line in the table shows where the decision to add or remove the next state of the hostile or friend rules would increase the total cost. Now that the baseline cost has been determined, the heuristic for finding the best combination of hostile, friend and non-declared rules can be utilized. This is accomplished by testing the removal of different states from both the hostile and friend rules for use in the non-declared rule. Hostile rule elements $h_{[j]}$ will be removed in sequence and cost will be calculated with element [j] being part of the non-declaration rule based upon their likelihood ratios. For simplicity, define a new index for rule elements based on their order statistic from the likelihood ratio calculations. Define $H, F,$ and ND as the hostile, friend and non-declared rules respectively. Let $h_{[k]}$ be the k^{th} ordered element of the hostile rule H where TCM is at a minimum and $f_{[m]}$ be the $(k+1)^{st}$ ordered element of F based on the order statistic index where $[k] = \arg\min_{[j]\in H}(LR^i_{[j]})$ and $[m] = \arg\max_{[j]\in F}(LR^i_{[j]})$. The element $h_{[j]}$ associated with the

smallest likelihood ratio will be the first candidate for removal; likewise, the first friend element tested is based on the element $f_{[j]} = 1$ associated with the largest likelihood ratio. Let C_{Ti} represent the temporary total cost of misclassification i calculated using equation (3-2). The ordered element contributing the largest decrease in total cost will be included in the non-declared rule. This process will be iterated until either $nd_{[j]} = 1$ for all [j] or the addition of another element to the non-declared rule would increase the minimum total cost. From the example in Table 3-9, the first hostile element to be tested is $h_{[3]}$ where $h_{[3]} = 0$ and $nd_{[3]} = 1$; the first friend element for comparison is $f_{[4]}$ where $f_{[4]} = 0$ and $nd_{[4]} = 1$. The total cost function will decide whether either state should remain part of the non-declared rule or not. The cost calculations in the algorithm that follows are based upon equation (3-2). The non-declared rule is incremented as follows:

1. Create non-declared rule initialized to zero, $ND = [0,0,...,0]^T$
2. Find $C_{T,LR}$ from equation (3-2)
3. Test $h_{[k]}$ element of the hostile rule versus f_m element of the friend rule
    a. Set $h_{[k]} = 0$, $nd_{[k]} = 1$
    b. Calculate C_{T1} from equation (3-2)
    c. Set $h_{[k]} = 1$, $nd_{[k]} = 0$, $f_{[m]} = 0$, $nd_{[m]} = 1$
    d. Calculate C_{T2} from equation (3-2)
    e. Set $nd_{[m]} = 0$, $f_{[m]} = 1$
4. If $\min(C_{T1},C_{T2}) < C_{T,LR}$
    a. If $C_{T1} \leq C_{T2}$

i. Set $nd_{[k]} = 1$, $h_{[k]} = 0$, $C_{T,LR} = C_{T1}$, $C_{T1} = 0$, $[k] = [k] – 1$

ii. Return to step 3

5. Else if $C_{T2} < C_{T1}$

a. Set $nd_{[m]} = 1$, $f_{[m]} = 0$, $C_{T,LR} = C_{T2}$, $C_{T2} = 0$, $[m] = [m] + 1$

6. Return to step 3

Else if $C_{T,LR} < \min(C_{T1}, C_{T2})$

If any $P(S_j \mid H) = P(S_j \mid F) = 0$, $nd_j = 1, h_j = 1, f_j = 1$

End

Once the above algorithm is completed, the expected minimum cost rule set for the specified grid point is created. The total cost found from this method will be denoted $C_{T,LR}$. This method could be implemented to the non-declaration heuristic in Figure 3-2 starting at step 2 to find the global minimum cost rules. The flowchart in Figure 3-4 further encapsulates the likelihood ratio process for a given grid point.

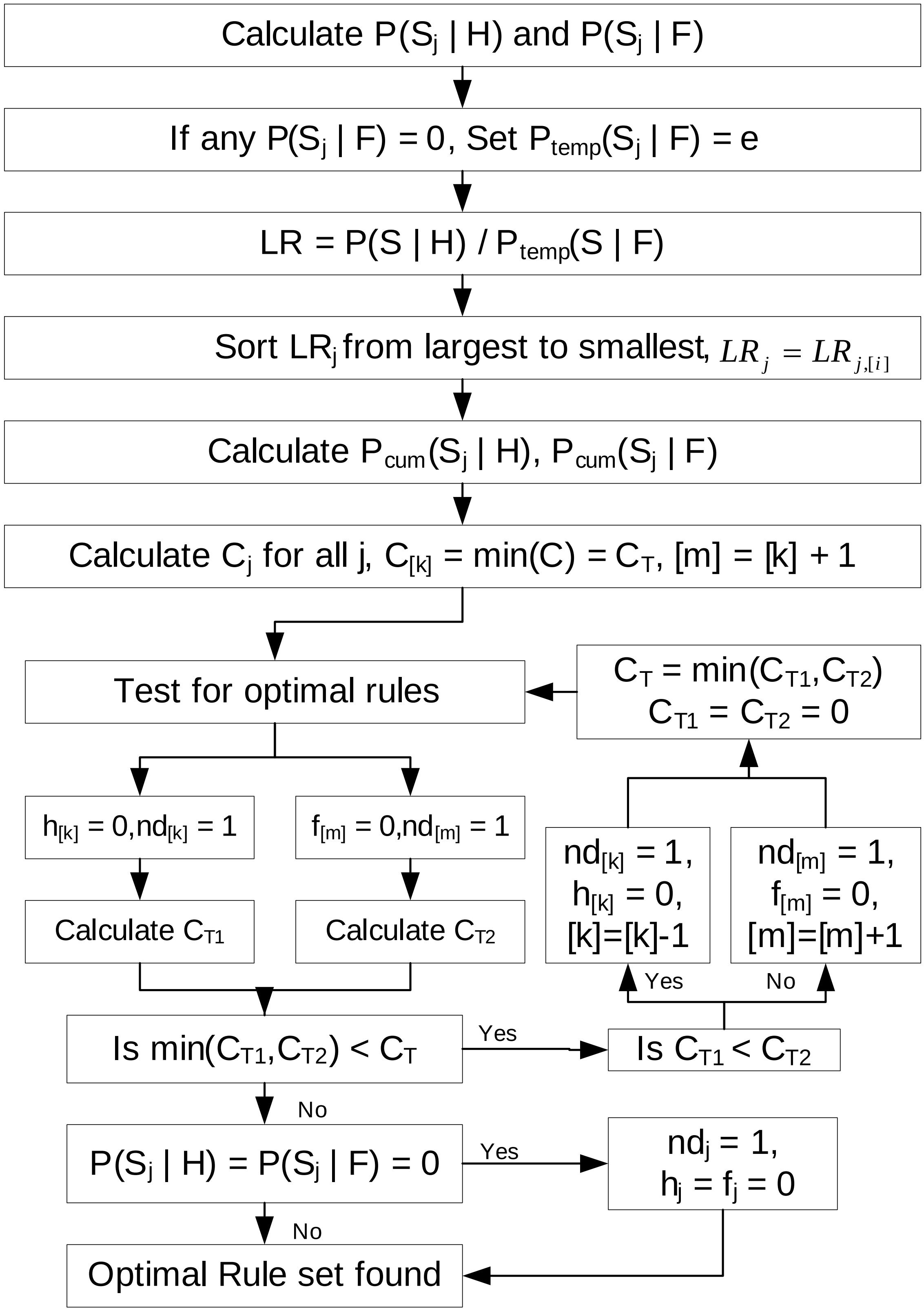


Figure 3-4: ISOC Fusion rule selection through likelihood ratios.

3.2.5 RSM Cost experiment

It became readily apparent that there were some subjective areas of this research that could affect optimal rules and costs. This subjective nature was tested through the use of a response surface methodology experiment. The costs of misclassification were based on a best guess, but changes could drastically affect the outcome of the different experiments. These costs and their interactions with correlation, both intra-correlation and inter-correlation, were considered. The sample size was held constant at 250 exemplars per class. In addition, the grid points tested were $\delta_j^i = (i-1) \times 0.1$ for $i = 1,2$ and $j = 1,...,6$. This was considered enough grid points to get a clear picture of the sample space. The optimal rules for each grid point were calculated using the likelihood ratio heuristic (IND_{LR}); these rules remained relatively constant across all grid points. The problem was designed as a full factorial with five factors. The ranges for the variables are as follows: $\rho_{\text{inter}} \in \{0,0.8\}$, $\rho_{\text{intra}} \in \{0,0.9\}$, $C_{FP} \in \{10,20\}$, $C_{FN} \in \{5,9\}$, $C_{ND} \in \{1,4\}$. The ranges for the costs were designed to ensure the following generally accepted inequality: $C_{FP} > C_{FN} >> C_{ND}$. The experimental response variable was $C_{T,LR}$.

3.2.6 PNN Non-Declaration Fusion Method (NF_{ND})

The PNN fusion method takes the posteriors output from the classifiers and uses them as features. One-third of the posterior probabilities are fed to the PNN from the test set to train the network. The validation set plus the remaining two-thirds of the test set exemplars are used for validation. The spread parameter was tested to achieve the highest possible accuracy. The outputs from the PNN were then put through an additional radial basis function and turned into posterior probabilities using Baye's rule. Decisions for classifications were once again made based on $T \pm \delta^i$ although there is

now only one indifference window $(\delta^i = (i-1)\times 0.5$ for $i = 1,...,11)$ under consideration.

As stated earlier, if $\begin{cases} P_k > T + \delta^i, \text{then the exemplar is labeled "}H\text{"} \\ P_k < T - \delta^i, \text{then the exemplar is labeled "}F\text{"} \\ \text{else the exemplar is labeled "ND"} \end{cases}$ for all i and k. Some validation set exemplars were such outliers that their associated PNN activations were zero. These were automatically assigned equal posterior probabilities of class membership, $P_k = 0.5$. When $i = 1$, $\delta^i = 0$ and these exemplars were forced to be “F” classifications since the cost of a false negative was considered to be more acceptable than the cost of a false positive. When $i > 1$, $\delta^i > 0$ and these exemplars with equal posterior probabilities fell within the indifference window and classified as “ND”. Note that this method allows “H”, “F” and “ND” indications. Figure 3-5 shows the PNN non-declaration classifier fusion process NF_{ND} as used in this research.

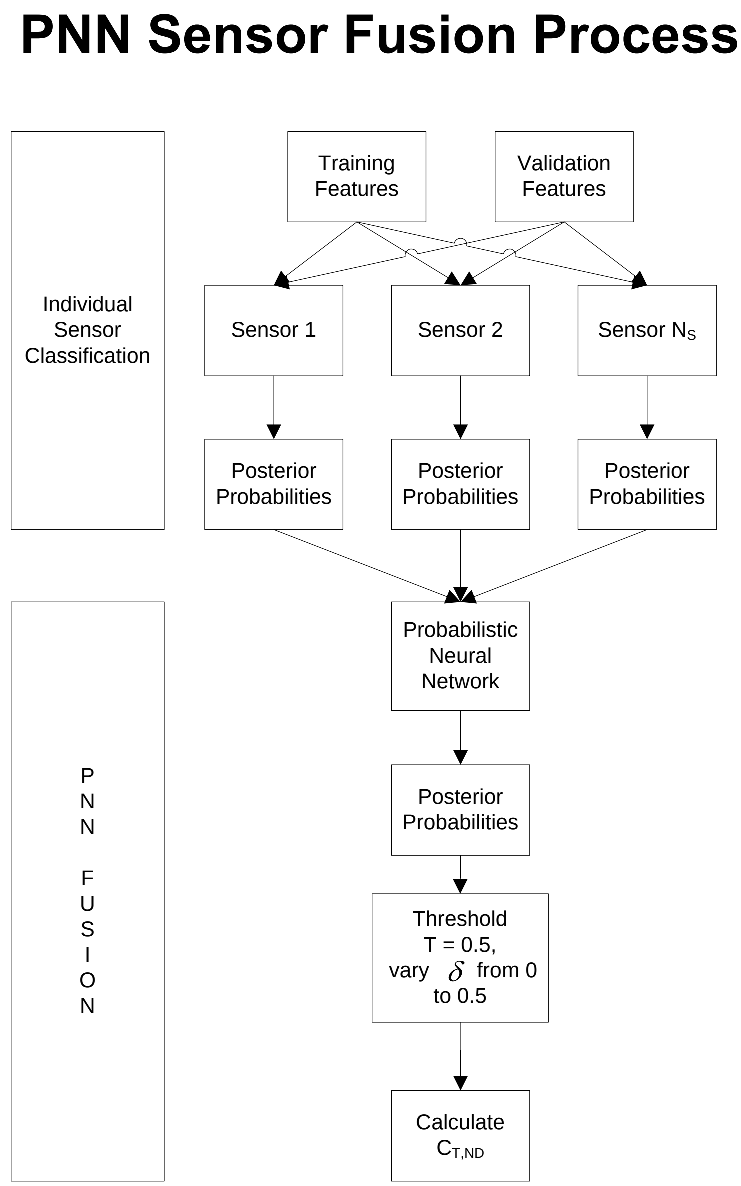


Figure 3-5: PNN Classifier Fusion Process (NF_{ND}).

The total cost of misclassification is again used as a cost function and allows a comparison between methods considered in this research. Total costs are compared in the results section to find optimal indifference windows for particular runs. Accuracy and the percent of non-declared will also be addressed.

3.3 Summary

This chapter introduced the framework for this research. The experimental design was explained as laid out in Leap (2004). The forced decision ISOC method was introduced as well as a heuristic to adapt the forced decision ISOC to incorporate a non-declared rule. The ISOC method using likelihood ratios was described. Next, the RSM experiment considering cost ranges and their interaction with correlation levels on TCM was introduced. Finally, the PNN fusion method was discussed. Table 3-10 shows the methods introduced in this chapter with their associated acronyms and cost labels.

Table 3-10: Acronyms and Costs.

Method	Acronym	Cost
ISOC	ISOC	$C_{T,I}$
ISOC Forced Decision	IFD	$C_{T,FD}$
ISOC Non-Declarations "Political Correctness"	IND_{PC}	$C_{T,PC}$
ISOC Non-Declarations Likelihood Ratio	IND_{LR}	$C_{T,LR}$
PNN Non-Declarations	NF_{ND}	$C_{T,NND}$

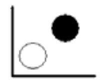
4. Findings and Analysis

4.1 Introduction

In this chapter, the three ISOC fusion heuristics developed in Chapter 3 were compared and contrasted to the PNN fusion method by their classification accuracies and costs found by executing the problems described in Table 2-6. Findings are also introduced for the RSM study investigating the interactions between cost and correlation.

4.2 General Findings

After a thorough investigation of the different methods, some common results were found. The first was that there was a consistent relationship between the total costs of misclassification achieved. It will be shown later in this section that $C_{T,FD} > C_{T,PC} \geq C_{T,NND}$ in a statistically significant sense where $C_{T,FD}$ is the TCM for the ISOC forced decision method (IFD), $C_{T,PC}$ is the TCM associated with the ISOC non-declaration political correctness heuristic (IND_{PC}), and $C_{T,NND}$ represents the TCM for the PNN non-declarations fusion method (NF_{ND}). The second result displayed the difference between method assumptions; the ISOC methods assume independence and do not react much to correlation while neural networks such as PNN fusion methods do not assume independence and react accordingly. This further supports the research of Storm *et al.,* (2003) and Leap *et al.,* (2004). It was discovered in problems 5 and 6 that the classifiers chosen were unable to adequately label exemplars creating disconcerting findings. The next sections of this chapter will step through each problem individually and further explore the analysis performed for this research effort.

4.3 Problem 1 Results: 4 Feature Case 

This data set was applied to the fusion processes described in Chapter 3. Problem 1 costs were consistent with the above inequality $C_{T,FD} > C_{T,PC} \geq C_{T,NND}$. Problem 1 demonstrated the IND_{PC} ability to reach a common optimum hostile rule as sample size was increased for a set correlation ρ. Table 4-1 shows the optimal ISOC non-declaration political correctness heuristic optimal hostile rules for a set sample size of 25 exemplars per class and $\rho = 0$. The indications are from classifier 1 (linear discriminant function) and classifier 2 (quadratic discriminant function) (L-Q), respectively. The run denotes the different random seeds used.

Table 4-1: IND_{PC} Optimal Hostile Rules (Sample Size = 25, ρ = 0).

	Indications (Linear - Quadratic)								
Run	**H-H**	**H-U**	**H-F**	**U-H**	**U-U**	**U-F**	**F-H**	**F-U**	**F-F**
1	1	1	0	0	0	0	0	0	0
2	1	1	0	0	0	0	1	0	0
3	1	0	1	1	0	0	0	0	0
4	1	1	0	0	0	0	0	0	0
5	1	1	1	0	0	0	1	0	0

The hostile rules above are variable across random number streams. As sample size is increased, the optimal rule set becomes constant. In fact, using a sample size of 500 exemplars per class, the optimal hostile rules for IND_{PC} are consistent only declaring a target as hostile if both classifiers indicate that it is hostile. Table 4-2 compares the optimal hostile rules over 5 random seeds with a sample size of 500 exemplars per class.

Table 4-2: IND_{PC} Optimal Hostile Rules (Sample Size = 500, ρ = 0).

	Indications (Linear - Quadratic)								
Run	H-H	H-U	H-F	U-H	U-U	U-F	F-H	F-U	F-F
1	1	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0
4	1	0	0	0	0	0	0	0	0
5	1	0	0	0	0	0	0	0	0

The same results were noticed when varying sample size for $\rho = 0.9$. Table 4-3 shows the variable hostile rules found for a sample size of 25 exemplars per class.

Table 4-3: IND_{PC} Optimal Hostile Rules (Sample Size = 25, ρ = 0.9).

	Indications (Linear - Quadratic)								
Run	H-H	H-U	H-F	U-H	U-U	U-F	F-H	F-U	F-F
1	1	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	1	0	0
3	1	0	0	1	0	0	1	0	0
4	1	1	0	0	0	0	1	0	0
5	0	1	0	1	0	0	1	0	0

Some of the most obvious states are confused with the addition of inter-correlation. The output state when both classifiers declare a target as hostile is not included in some of the optimal hostile rule. Once sample size is sufficiently increased to 500 exemplars per class, the rules stabilize to two common optimal hostile rules as shown in Table 4-4.

Table 4-4: IND_{PC} Optimal Hostile Rules (Sample Size = 500, = 0.9).

	Indications (Linear - Quadratic)								
Run	**H-H**	**H-U**	**H-F**	**U-H**	**U-U**	**U-F**	**F-H**	**F-U**	**F-F**
1	1	0	0	0	0	0	1	0	0
2	1	0	0	0	0	0	1	0	0
3	1	0	0	0	0	0	0	0	0
4	1	0	0	0	0	0	1	0	0
5	1	0	0	0	0	0	0	0	0

Thus, it can be inferred that as sample size increases, the optimal hostile rule becomes more invariable for IND_{PC}. There were some volatile states, but overall the rule steadied itself. The other problems also demonstrated this characteristic.

The indifference windows became more stable as sample size increased in this problem. This is a useful result to locate optimal grid point settings for the chosen rule set. Figure 4-1 shows a histogram of δ_1 (indifference window on classifier 1) varied from sample sizes of 25 to 500. The optimal indifference window falls around 0.3 with high sample size. The histogram of sample size 25 is much more difficult to determine an optimal grid point setting.

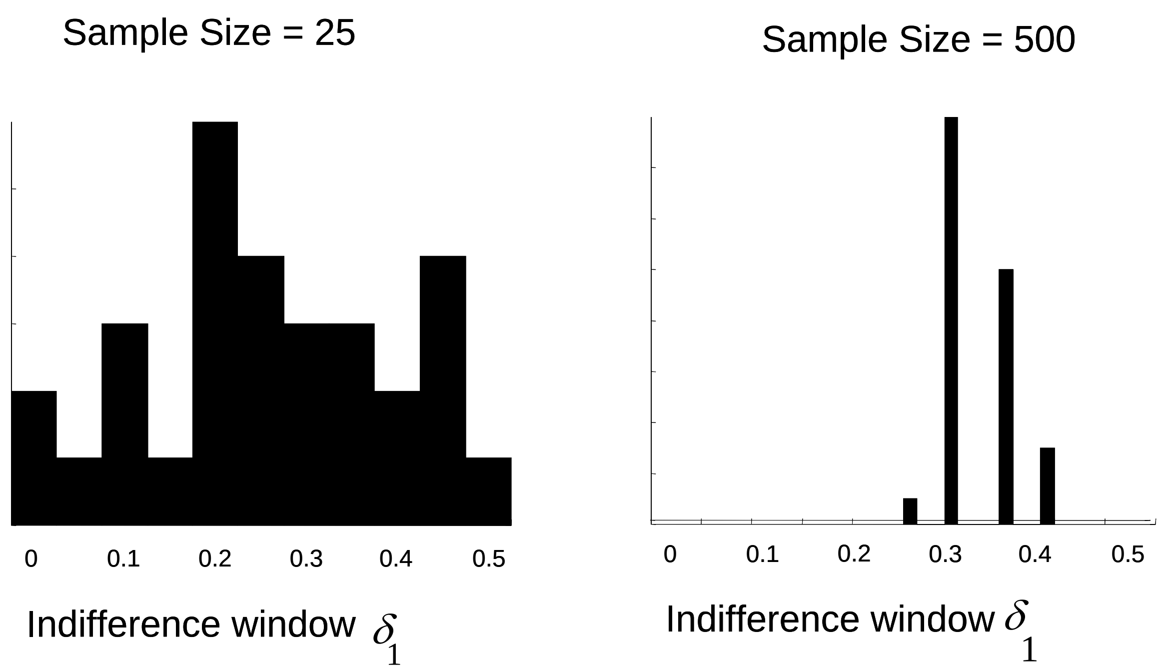


Figure 4-1: Indifference Window Comparison.

This problem was relatively easily separable and as a result, the IND_{PC} (ISOC non-declaration political correctness heuristic) was comparable to the PNN fusion NF_{ND} and there was no statistical difference between the two methods. Table 4-5 shows the paired t-test conducted on the difference in mean costs between the different methods while varying sample size and ρ. $C_{T,FD}$ represents the costs achieved through the ISOC forced decision heuristic; $C_{T,PC}$ represents the costs from ISOC non-declaration political correctness heuristic and $C_{T,NND}$ accounts for the total cost of misclassification found through PNN fusion allowing non-declarations. The highlighted rows in Table 4-5 represent the only difference in mean costs that were not rejected. All other rows in the table were statistically different at significance level $\alpha = 0.05$.

Table 4-5: Paired t-tests of Difference in Mean Costs.

Methods	Sample Size	p	p-value	lower CI	upper CI	t-statistic	t-critical
$C_{T,PC}$ – $C_{T,NND}$	All	All	0.006	0.0073	Inf	2.5153	± 1.9622
$C_{T,FD}$-$C_{T,PC}$	25	0.0	0	0.4	0.9	5.9	± 2.0017
$C_{T,FD}$-$C_{T,PC}$	25	0.9	0	0.4	1.0	4.4	± 2.0017
$C_{T,FD}$-$C_{T,NND}$	25	0.0	0	0.4	0.9	5.9	± 2.0017
$C_{T,FD}$-$C_{T,NND}$	25	0.9	0	0.5	1.2	5.1	± 2.0017
$C_{T,PC}$-$C_{T,NND}$	**25**	**0.0**	**0.6**	**-0.1**	**0.2**	**0.5**	**± 2.0017**
$C_{T,PC}$-$C_{T,NND}$	**25**	**0.9**	**0.1**	**0**	**0.3**	**1.6**	**± 2.0017**
$C_{T,FD}$-$C_{T,PC}$	500	0.0	0	0.5	0.6	14.4	± 2.0017
$C_{T,FD}$-$C_{T,PC}$	500	0.9	0	0.6	0.9	12.0	± 2.0017
$C_{T,FD}$-$C_{T,NND}$	500	0.0	0	0.5	0.6	14.4	± 2.0017
$C_{T,FD}$-$C_{T,NND}$	500	0.9	0	0.6	0.9	11.4	± 2.0017
$C_{T,PC}$-$C_{T,NND}$	500	0.0	0	0.1	0.1	8.7	± 2.0017
$C_{T,PC}$-$C_{T,NND}$	500	0.9	0	-0.1	0.0	-2.4	± 2.0017

Note that IND_{PC} was statistically less than NF_{ND} for sample size of 500 and ρ = 0.9 as shown in the last row of the table although this is a small difference in a practical sense. There were instances such as this when the ISOC heuristics were able to outperform the neural networks, but in general the opposite remained true.

Cost and accuracy were compared through the use of a parametric analysis. The parametric analysis compared average costs and accuracies across all sample sizes for ISOC, IFD, IND_{PC}, NF, and NF_{ND}. Since lower costs were preferred, the average costs were transformed by $C_i = \dfrac{(1-\overline{C_i})}{\max_i(C_i)}$. Each method generates a score through the following equation: $(1-\alpha)\times C_i + \alpha \times A_i$ where A_i represents the average accuracy for method i and α is varied along the x-axis to compare the methods. When α = 0, cost is the most important factor and when α = 1, accuracy becomes the most important. Otherwise, α represents the weighting for the two factors being compared. Figure 4-2 displays these average scores varied across α for no correlation.

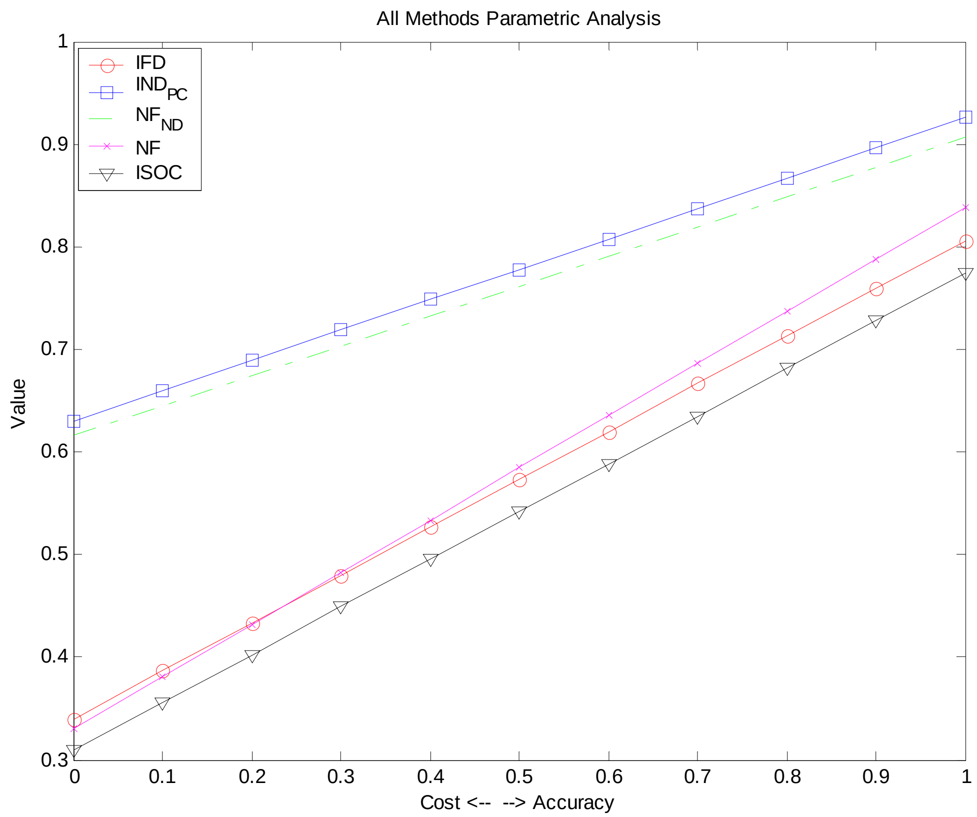


Figure 4-2: Parametric Analysis of Cost and Accuracy (ρ= 0.0).

IND_{PC} outperforms all other methods across all weightings α, although there is no statistical difference between IND_{PC} and NF_{ND}. The above analysis varies when correlation increases. Figure 4-3 shows the five methods with high levels of correlation.

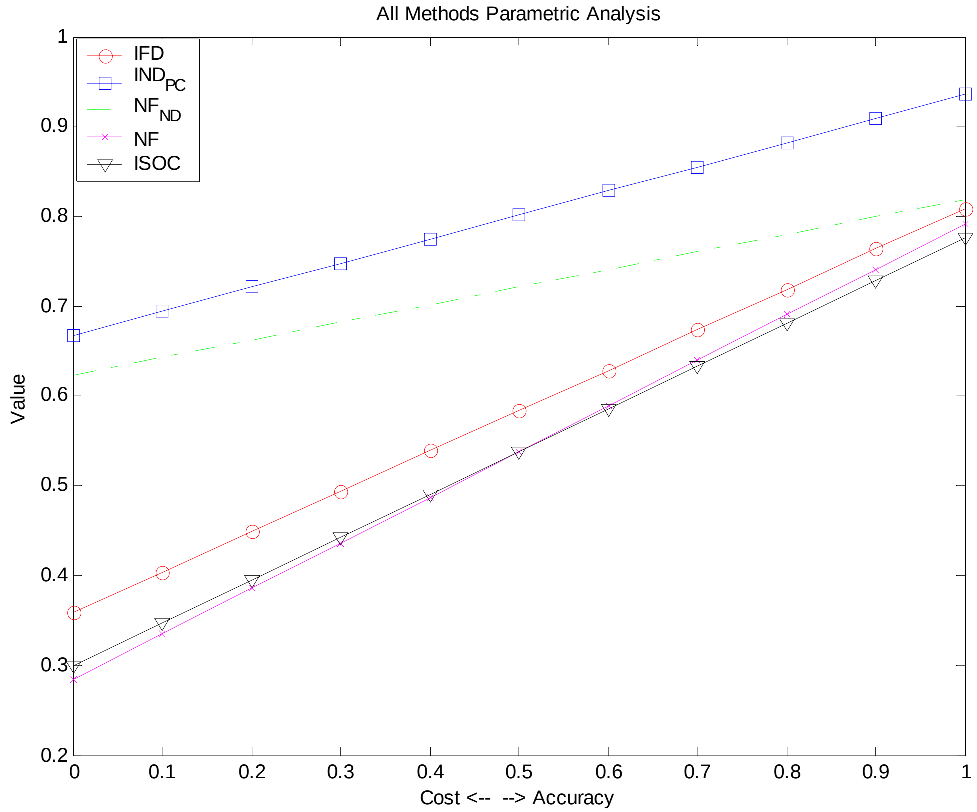


Figure 4-3: Parametric Analysis of Cost and Accuracy (ρ = 0.9).

The above parametric analysis shows that the PNN fusion methods react to changes in levels of correlation where the ISOC fusion methods do not. IND_{PC} is now statistically better than all other methods across all weightings of cost and accuracy for problem 1.

4.4 Problem 2 Results: 8 Feature Case 

This data set was applied to the fusion processes described in chapter 3. This problem will be used to show the cost inequality described in the general findings section above ($C_{T,FD} > C_{T,PC} \geq C_{T,NND}$). Table 4-6 shows the costs as sample size and correlation were varied.

Table 4-6: ISOC Heuristics Vs. NF_{ND} Costs.

Run	Sample Size	ρ	$C_{T,FD}$	$C_{T,PC}$	$C_{T,NND}$
1	25	0.00	1.28	0.74	0.75
1	25	0.20	1.53	0.68	0.69
1	25	0.40	2.50	0.92	0.35
1	25	0.60	1.96	1.00	0.67
1	25	0.80	2.32	1.00	0.77
1	25	0.90	2.26	0.90	1.00
1	50	0.00	1.83	0.78	0.73
1	50	0.20	1.89	0.95	0.60
1	50	0.40	2.00	0.92	0.69
1	50	0.60	2.50	0.94	0.72
1	50	0.80	2.24	0.90	0.90
1	50	0.90	2.25	0.90	0.87
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
5	500	0.80	2.28	0.97	0.78
5	500	0.90	2.24	0.93	0.87
5	1000	0.00	1.88	0.95	0.66
5	1000	0.20	1.96	0.94	0.67
5	1000	0.40	2.15	0.94	0.75
5	1000	0.60	2.07	0.93	0.78
5	1000	0.80	2.20	0.93	0.81
5	1000	0.90	2.33	0.95	0.83

Notice that the above inequality holds for most instances although there are a few times when $C_{T,PC} < C_{T,NND}$ in the table. Further inspection of the costs through the use of paired t-tests establishes that the means for the outputs of the three classifiers are statistically different with $C_{T,PC} \geq C_{T,NND}$ holding as stated above. Table 4-7 displays the results of the two-tailed paired t-tests conducted on $C_{T,PC}$ and $C_{T,NND}$. Sample size was varied across all sample sizes tested, but ρ was displayed at the extremes $\rho = \{0, 0.9\}$. The null hypothesis tested is that the means are equal. The p-value displays the probability of observing the given result by chance given that the null hypothesis is true; p-values less than $\alpha = 0.05$ indicate no statistical difference in means. The high and low

CI's in the table denote the confidence interval on the mean of $C_{T,PC} - C_{T,NND}$. A confidence interval that contains zero represents an observed result that fails to reject the null hypothesis. Finally, the t statistic and t critical value compare the difference in means. A t statistic that is greater than the t critical results in rejecting the null hypothesis that there is no difference between the means.

Table 4-7: Paired T-Test of $C_{T,NND}$ and $C_{T,PC}$ Problem 2 Costs.

Sample Size	ρ	p-value	Low CI	High CI	t-statistic	t-critical
25	**0.0**	**0.12**	**-0.05**	**0.37**	**1.75**	**± 2.011**
50	0.0	0.03	0.03	0.34	2.72	± 2.011
100	0.0	0.00	0.22	0.30	14.42	± 2.011
250	0.0	0.00	0.17	0.32	7.70	± 2.011
500	0.0	0.00	0.25	0.32	19.37	± 2.011
1000	0.0	0.00	0.27	0.34	18.29	± 2.011
25	0.9	0.05	0.00	0.54	2.33	± 2.011
50	0.9	0.02	0.03	0.19	3.06	± 2.011
100	0.9	0.00	0.10	0.22	6.35	± 2.011
250	0.9	0.00	0.07	0.17	5.20	± 2.011
500	0.9	0.01	0.03	0.14	3.49	± 2.011
1000	0.9	0.00	0.10	0.18	7.71	± 2.011

Only one instance in the table fails to reject the null at sample size = 25 and ρ = 0. All of the other observations show a statistical difference between the mean of $C_{T,NND}$ and $C_{T,PC}$. A paired t-test of all of the observations from problem 2 also shows that the means for all three costs are statistically different as shown in Table 4-7. Thus, it can be inferred that the cost relationship holds under differing levels of correlation. A graphical representation of the results from the paired t-tests can be seen in Figure 4-4, Figure 4-5, and Figure 4-6. ISOC non-declaration political correctness heuristic costs (IND_{PC}), ISOC forced decision costs (IFD) and PNN fusion (NF_{ND}) costs are paired across sample sizes

and levels of correlation ρ. Clusters of points above the 45^{o} line indicate that Y values dominate X; X values dominate Y when the majority of the paired observations fall below the 45^{o} line. Figure 4-4 graphically depicts the costs relationship between PNN fusion allowing non-declarations and the ISOC non-declaration political correctness heuristic.

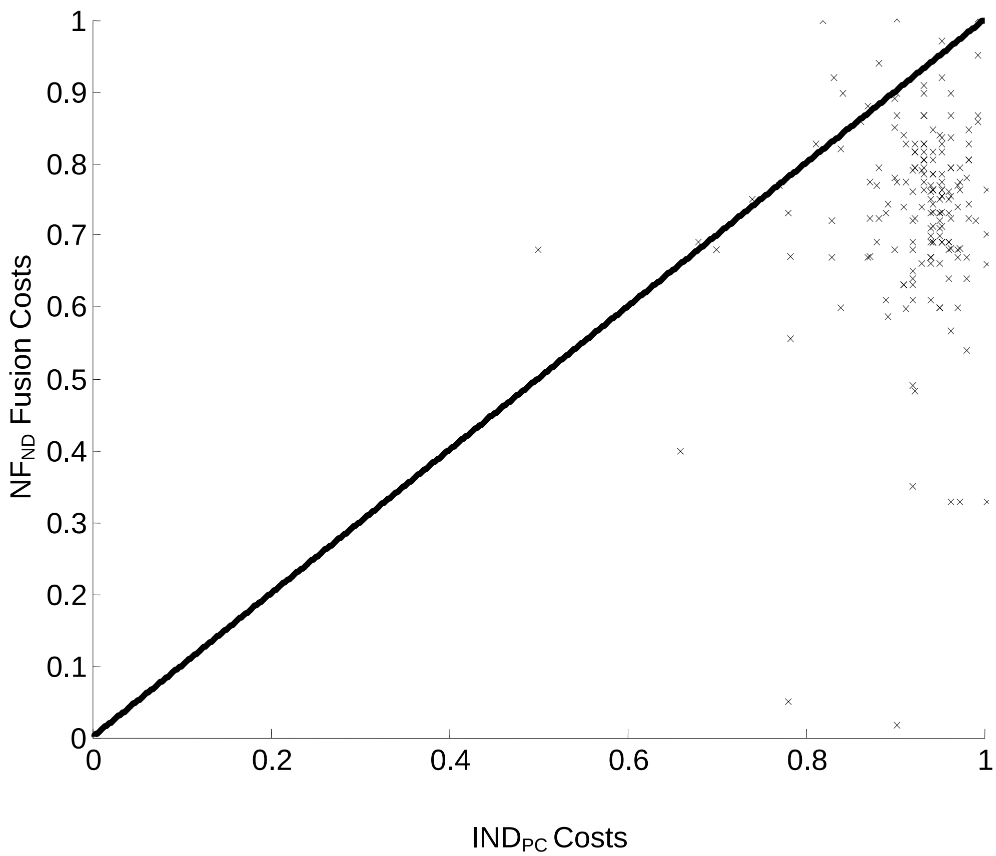


Figure 4-4: Problem 2 IND_{PC} Heuristic and NF_{ND} Fusion Ordered Pairs.

The ISOC non-declaration political correctness heuristic costs are always greater than the PNN fusion allowing non-declarations. The centroid of the cluster of data falls at $(0.92,0.72)$, which represents the paired means of the data. Continuing with this

graphical approach in Figure 4-5, it is clear that the ISOC forced decision heuristic falls short of the other two methods because its costs completely dominate the other approaches. Figure 4-5 shows ISOC forced decision paired with the ISOC non-declaration political correctness heuristic.

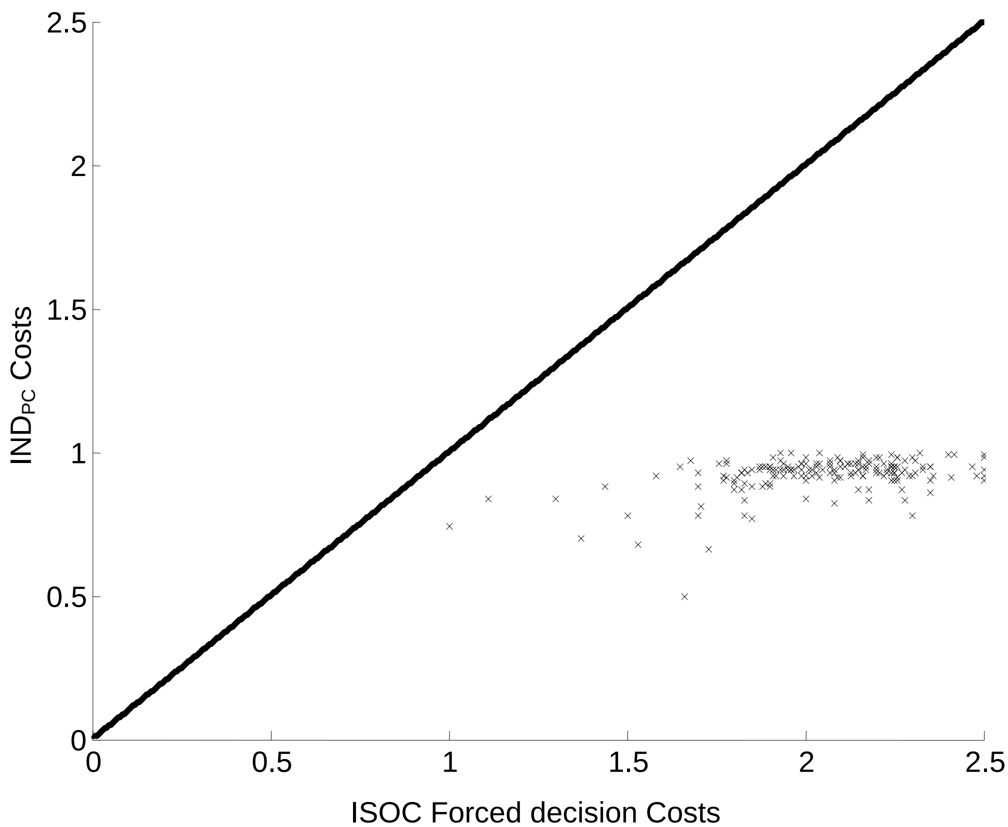


Figure 4-5: Problem 2 IFD and IND_{PC} Ordered Pairs.

The centroid for this pairing of data falls at $(2.05, 0.92)$ and it is clear that the ISOC forced decision costs are much higher than the ISOC non-declaration political correctness heuristic costs. Following the logic of the original inequality suggested, $C_{T,FD} > C_{T,PC} \geq C_{T,NND}$ and the plot of ISOC forced decision costs versus PNN

fusion costs should and does show the same story. ISOC forced decision is the most costly method of labeling targets as seen again in Figure 4-6.

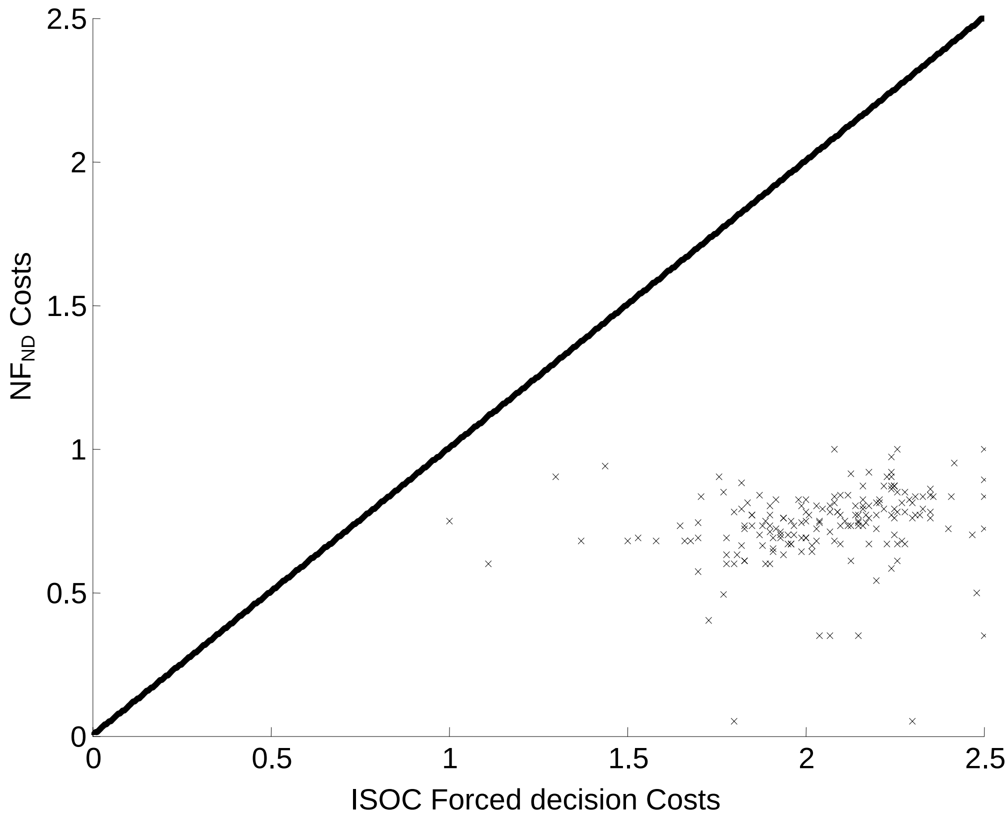


Figure 4-6: Problem 2 IFD and NF_{ND} Ordered Pairs.

The centroid for Figure 4-6 is located at $(2.05, 0.72)$ again showing the cost relationship holding true to the inequality.

Problem 2 also followed the pattern displayed for problem 1 of a global hostile rule being reached as sample size was increased. In fact, problem 2 reached a single hostile rule comprised of state (H-H) for a sample size of 500 exemplars per class regardless of the amount of correlation introduced.

4.5 Problem 3: 8 Feature with Autocorrelation Case

This data set was applied to the fusion processes described in Chapter 3. Problem 3 held true to the inequality described in the general findings section with $C_{T,FD} > C_{T,PC} \geq C_{T,NND}$. Figure 4-7 shows the average costs incurred from each of the three methods at varied levels of inter-correlation ρ and autocorrelation ρ_{auto}.

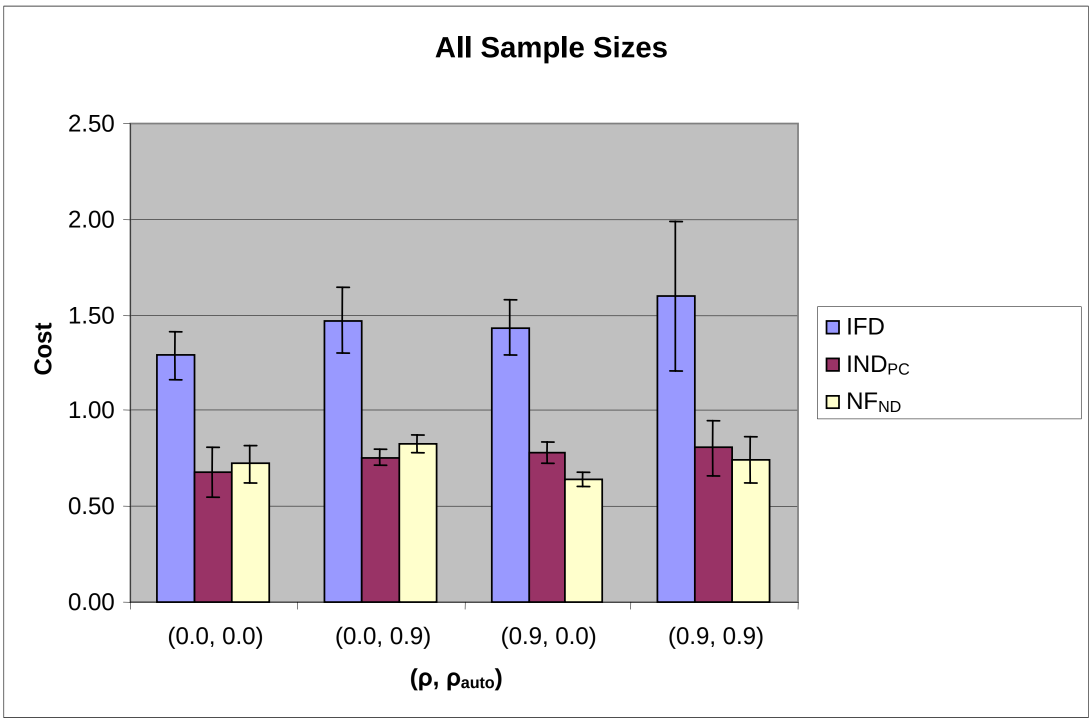


Figure 4-7: Cost Comparisons Between Methods.

The error bars on Figure 4-7 represent a 95% confidence interval of the true mean cost. Consider the ISOC forced decision costs which represent the largest costs in the above figure. The ISOC forced decision error bars all overlap showing that the average IFD costs are not statistically different. The error bars on ISOC forced decision are all

above the error bars for IND_{PC} and NF_{ND} fusion showing that ISOC forced decision is statistically more costly than IND_{PC} and NF_{ND} fusion. Next, consider the bars associated with IND_{PC}. There is no clear difference between any of the individual error bars. Thus, ISOC non-declarations political correctness heuristic is robust to correlation in this problem. Finally, consider the PNN allowing non-declarations error bars. There is a statistical difference between NF_{ND} fusion costs when ρ and ρ_{auto} (ρ, ρ_{auto}) are changed from $(0.9, 0.0)$ to $(0.0, 0.9)$. The NF_{ND} performed consistently with the exception of the improvement when autocorrelation was high. IND_{PC} and NF_{ND} were statistically different at $(0.0, 0.9)$ further showing that the NF_{ND} was able to perform well at this setting. Otherwise, there was no statistical difference between the two costs. Thus, the inequality stated above holds true ($C_{T,FD} > C_{T,PC} \geq C_{T,NND}$). When correlation levels were increased and sample size was set at 500 exemplars per class, the hostile rule had a tendency to include the output state $("F","H")$ as shown in Table 4-8. This could represent the fact that the correlation of the data is such that the quadratic function is able to discriminate between the classes more often.

Table 4-8: Optimal Hostile Rules for Sample Size = 500, ρ = 0.9.

	Indications (Linear - Quadratic)								
Run	**H-H**	**H-U**	**H-F**	**U-H**	**U-U**	**U-F**	**F-H**	**F-U**	**F-F**
1	1	0	0	0	0	0	1	0	0
2	1	0	0	0	0	0	1	0	0
3	1	0	0	0	0	0	0	0	0
4	1	0	0	0	0	0	1	0	0
5	1	0	0	0	0	0	0	0	0

The percent of non-declarations resulting from NF_{ND} were dependent upon the level of correlation. As correlation increased, the NF_{ND} indifference window also grew causing more non-declared exemplars; the result is the increase in costs shown in Figure 4-7. Figure 4-8 compares the optimal $\delta_{NF_{ND}}$ locations when $\rho = 0$ plotted versus when $\rho = 0.9$. A triangle in the plot represents the optimal indifference window size for the PNN with non-declarations paired between no correlation and 0.9 correlation; each triangle is an instance of sample size and random seed (5 sample sizes x 5 runs = 25 points). The size of the window is in direct relation to the percent of non-declarations. The figure shows that the PNN fusion method indifference window increases as correlation increases. Thus, the PNN reacted to increases in correlation while the ISOC fusion methods declarations were consistent; PNN fusion had to non-declare more exemplars as correlation levels were increased in reaction to the correlation and ISOC heuristics assumed independence and remained relatively unchanged.

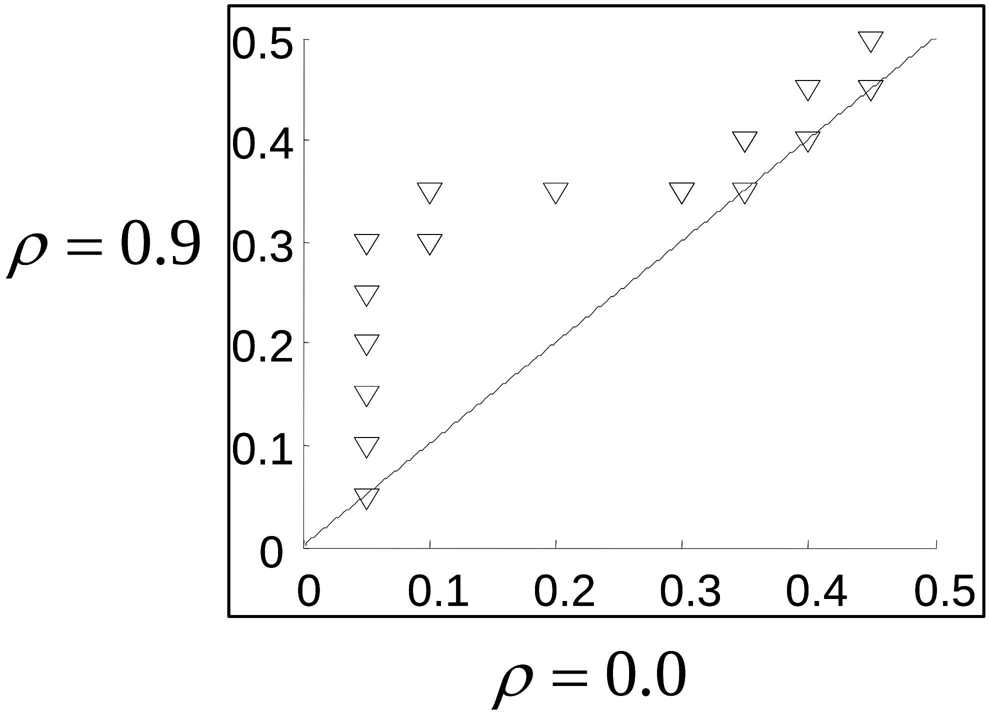


Figure 4-8: PNN Indifference Window Locations.

The IND_{PC} method's indifference window settings were relatively robust to the correlation. Figure 4-9 compares the optimal locations of the indifference window on the linear classifier while varying inter-correlation. The histograms are very similar further supporting the robustness of the ISOC fusion methods to correlation due to the assumption of independence.

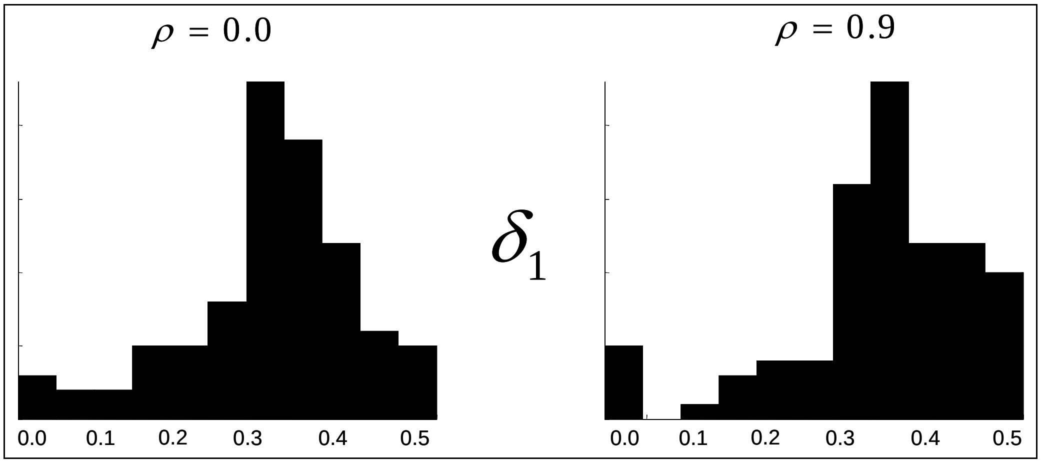


Figure 4-9: Linear Indifference Window varied by ρ .

This problem also displayed the PNN fusion with non-declaration method's decrease in accuracy associated with induced levels of correlation; this NF_{ND} "weakness" is consistent with the past research efforts of Storm *et al.,* (2003) and Leap *et al.,* (2004) showing the PNN reacts to levels of correlation. As the levels of correlation were increased, the NF_{ND} method's ability to classify incoming exemplars was hampered in some instances. The changes in correlation levels had less of an effect on the ISOC non-declaration political correctness heuristic since it assumes independent features. ISOC forced declarations had lower accuracies as expected because the method does not allow any fused non-declarations. IND_{PC} and NF_{ND} were both able to improve accuracy by ridding exemplars with high probabilities of misclassification. Figure 4-10 shows pair-wise accuracies for NF_{ND} at the extremes for correlation ($(\rho = 0, \rho_{auto} = 0)$ to $(\rho = 0.9, \rho_{auto} = 0.9)$).

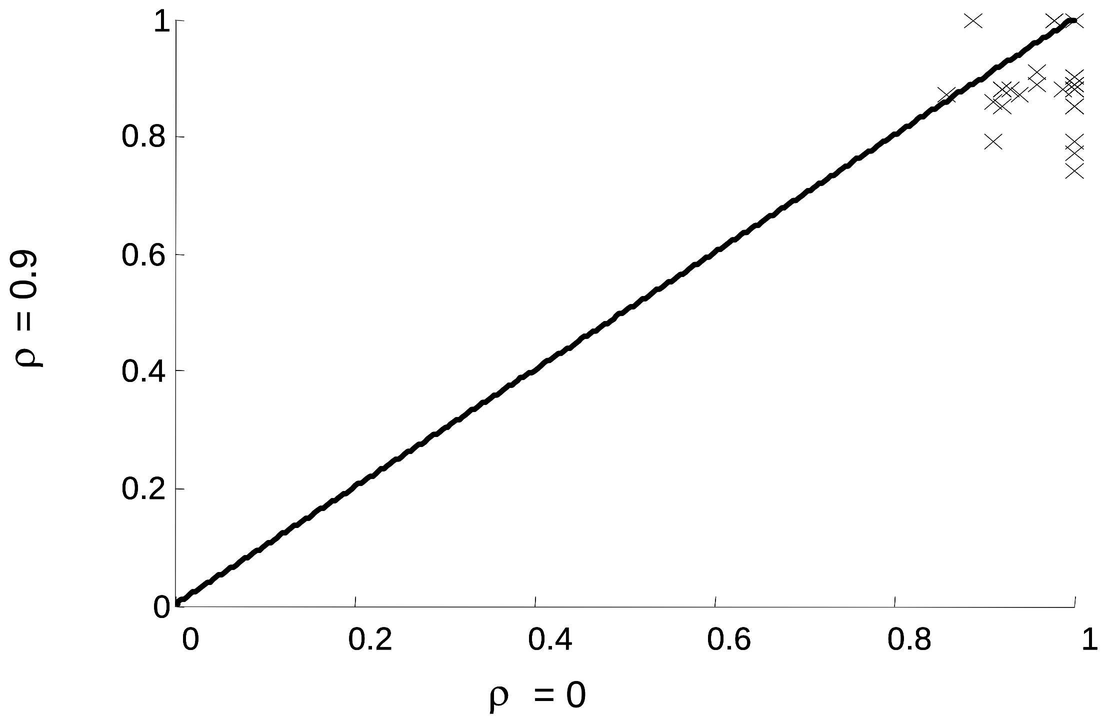


Figure 4-10: NF_{ND} Accuracy (No Correlation vs. High Correlation).

Further inspection through a paired t-test shows a statistical difference between the means for PNN fusion with and without correlation (both ρ and ρ_{auto}). Dividing the space from Figure 4-10 to consider all four extreme points for $\rho \in \{0,0.9\}$ and $\rho_{auto} \in \{0,0.9\}$ generated Figure 4-11.

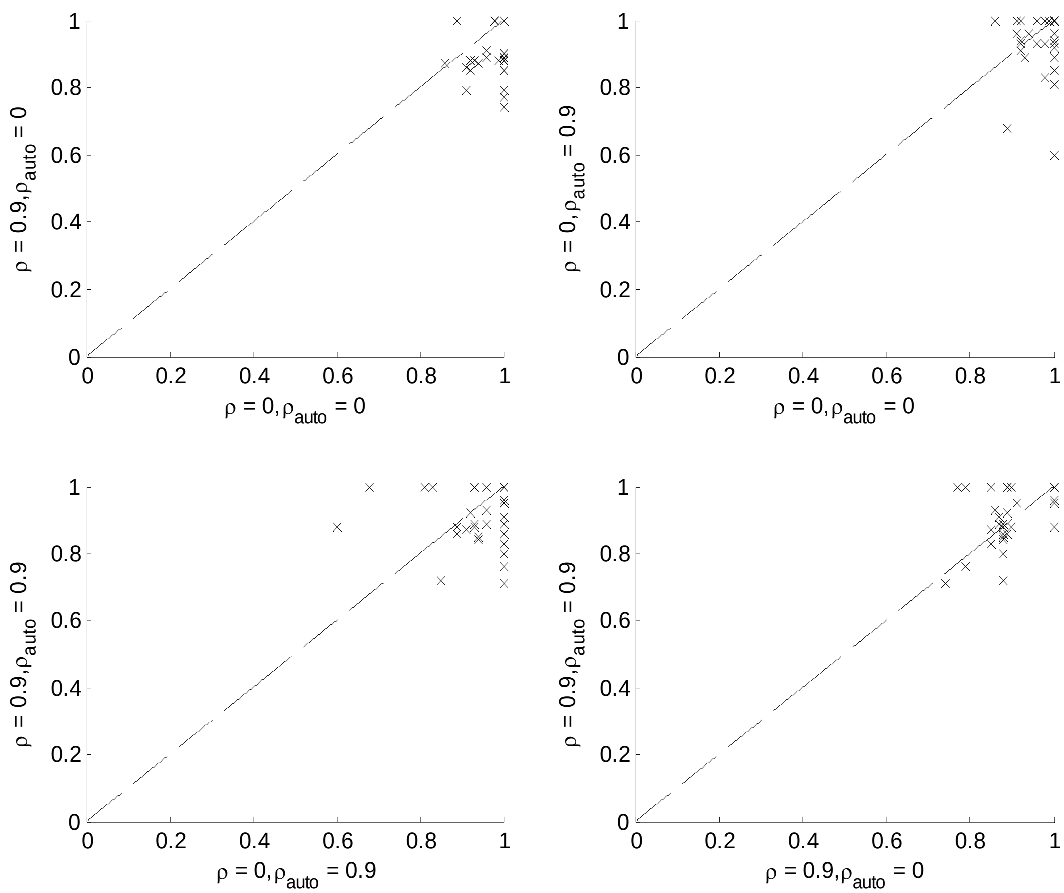


Figure 4-11: PNN Fusion Accuracy Pairwise Comparison of Correlation.

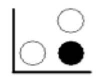
The difference is not as clear in the graphs in Figure 4-11 as it was in the cost comparisons, but two tailed paired t-tests of all four of the above plots proved that the only statistical difference in accuracies occurred when $\rho_{auto} = 0$ and ρ was varied between 0 and 0.9 (the upper left plot of Figure 4-11). The first row of Table 4-9 shows this occurrence with a p-value less than 0.05, a t-statistic greater than the t-critical value and a confidence interval which does not contain 0.

Table 4-9: PNN Fusion Accuracy Pair-wise T-test Results.

ρ_x	$\rho_{auto,\,x}$	ρ_y	$\rho_{auto,\,y}$	p-value	CI Low	CI High	t-statistic	t-critical
0.0	0.0	0.9	0.0	0.000	0.052	0.109	5.677	± 2.001
0.0	0.0	0.0	0.9	0.058	-0.001	0.075	1.933	± 2.001
0.0	0.9	0.9	0.9	0.240	-0.019	0.074	1.188	± 2.001
0.9	0.0	0.9	0.9	0.414	-0.055	0.023	-0.823	± 2.001

Thus, the PNN was susceptible to inter-correlation assuming that there was no autocorrelation present. Otherwise, the PNN accuracies remained statistically similar.

4.6 Problem 4: 8 Feature Triangle Case



This data set was applied to the fusion processes described in Chapter 3. This problem also held the same characteristics as discussed in the general findings section. The costs were found to follow the general inequality $C_{T,FD} > C_{T,PC} \geq C_{T,NND}$ in a statistical sense through paired t-tests. In addition, the optimal hostile rules followed the generally observed improvement as sample size increased as shown in problem 1. Further review of these hostile rules showed that the rules were more variable for high levels of correlation; the rules remained relatively constant when there was no correlation present for all sample sizes. Figure 4-12 shows a histogram comparing the optimal hostile rules as ρ is varied. Table 4-10 displays the numbered optimal hostile rules found using the ISOC non-declaration political correctness heuristic relating to Figure 4-12.

Table 4-10: Problem 4 Numbered Optimal Hostile Rules.

	Hostile Rules								
Rule #	H-H	H-U	H-F	U-H	U-U	U-F	F-H	F-U	F-F
2	1	0	0	0	0	0	0	0	0
3	0	1	0	0	0	0	0	0	0
11	1	1	0	0	0	0	0	0	0
13	1	0	0	1	0	0	0	0	0
16	1	0	0	0	0	0	1	0	0
48	1	1	0	1	0	0	0	0	0
51	1	1	0	0	0	0	1	0	0
54	1	0	1	1	0	0	0	0	0

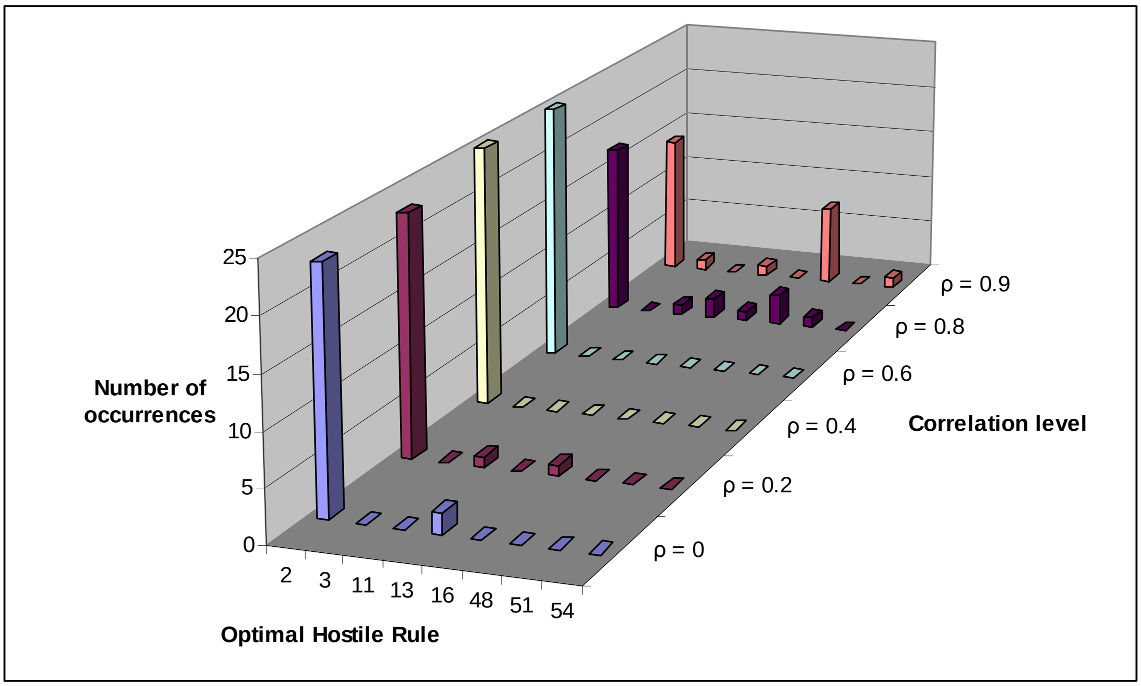
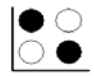


Figure 4-12: Problem 4 Optimal Hostile Rules Histogram.

The optimal rule set was constant for $\rho \in \{0.4, 0.6\}$ and very steady for $\rho \in \{0.0, 0.2\}$. This problem became more separable as correlation increased for both classifiers; the optimal rule went from only declaring an exemplar as hostile when both classifiers labeled it hostile to a more aggressive hostile rule declaring the exemplar

hostile when at least one of the classifiers labeled it hostile. The optimal hostile rules were very dependent on the level of correlation in this problem. With no correlation between the feature sets, IND only declared the output state (H-H) as hostile. As correlation increased, more states became part of the hostile rule.

4.7 Problem 5: 8 Feature XOR Case



This data set was applied to the fusion processes described in Chapter 3. The general results held for this problem as well. The ISOC non-declaration and HIND1 methods were unable to get adequate separation of the classes. The linear classifier is not apt for slicing an XOR problem such as this. Thus, the PNN outperformed all other methods. Figure 4-13 shows a pair-wise comparison of the costs further supporting the cost inequality stated earlier.

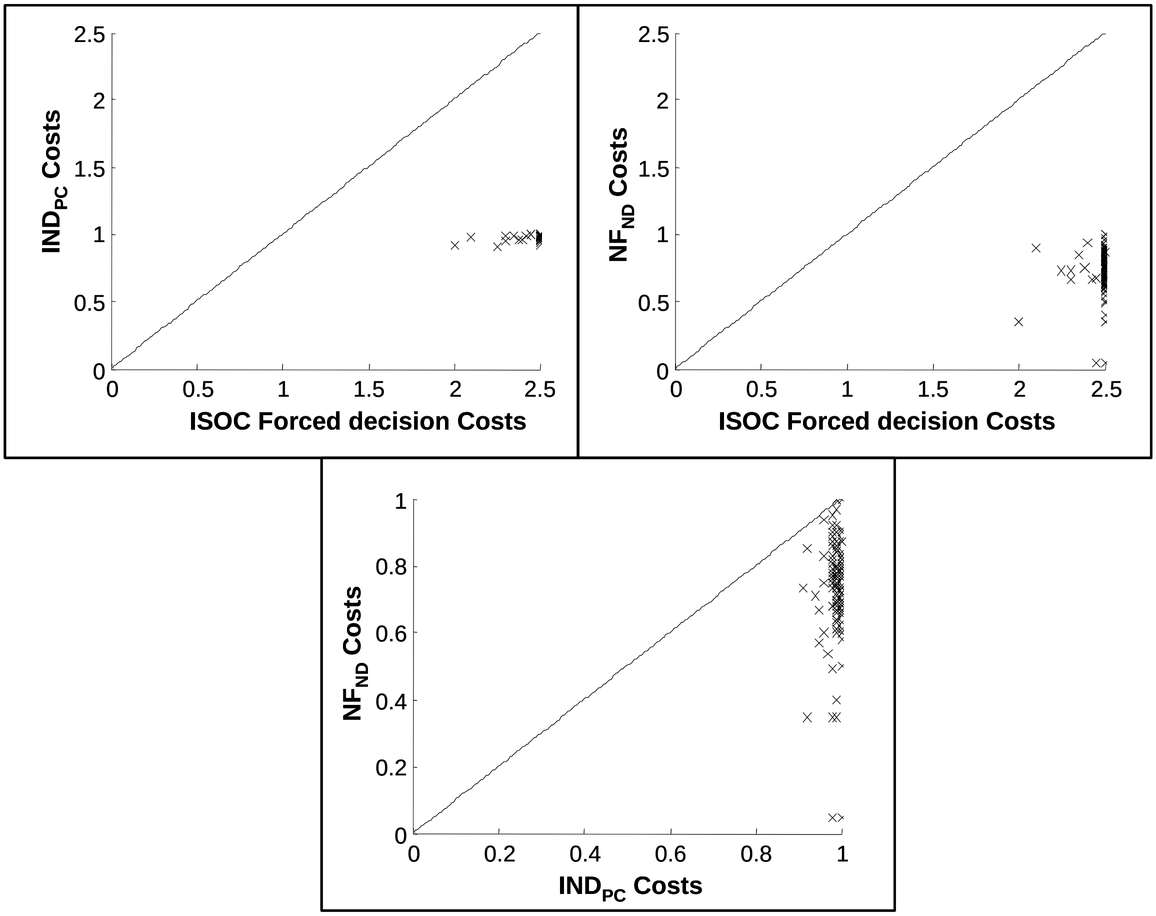


Figure 4-13: Marginal Costs.

Also, the box plots in Figure 4-14 show the same result. The box plots show that the average cost of IND_{PC} and NF_{ND} are different. The left box plot clearly shows that the lower 50th percentile of the difference between IND_{PC} and NF_{ND} is greater than zero. Paired t-tests also showed the inequality of $C_{T,FD} > C_{T,PC} \geq C_{T,NND}$ to hold true once again.

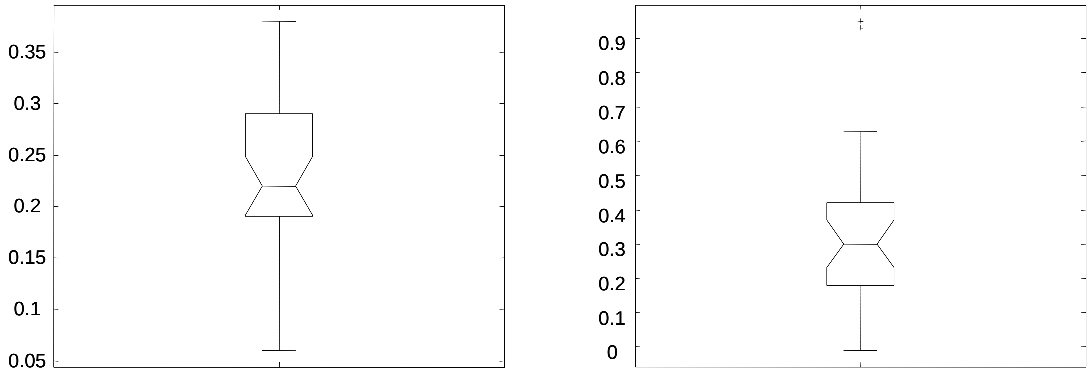


Figure 4-14: IND_{PC} vs NF_{ND} Fusion Box Plots.

With the linear classifier's inability to separate the groups, the linear indifference window returned unexpected results. When the sample size was increased from 25 to 500, the optimal delta window setting decreased. This is counterintuitive because the linear discriminant function is a poor classifier for an XOR problem and logic would suggest that the indifference window would be large allowing more individual non-declarations from the classifier. Figure 4-15 shows a histogram of the linear indifference window setting as sample size was varied from 25 to 500.

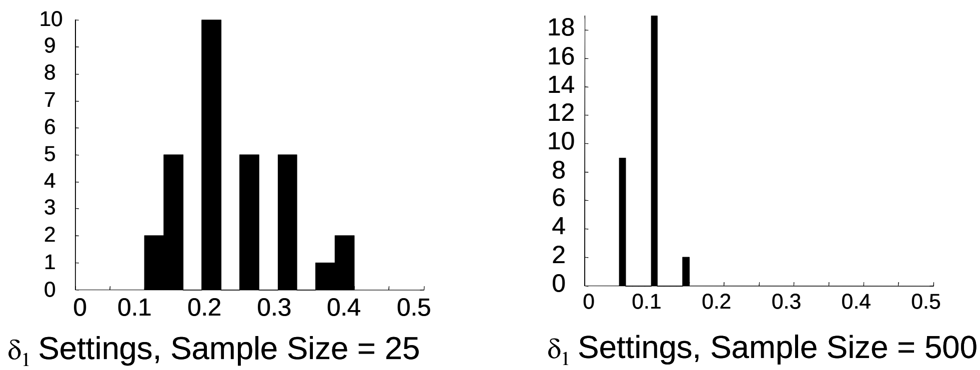
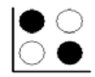


Figure 4-15: Linear Classifier Indifference Window Histogram.

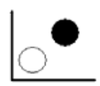
Due to the poor classification accuracy of the linear classifier, this problem and problem 6 were both reaccomplished using a PNN as classifier 1 and the quadratic discriminant function as classifier 2. Unfortunately, the PNN was unable to improve performance by acting as an individual classifier. The class means were too close to allow for any true separation.

4.8 Problem 6: 8 Feature XOR with Autocorrelation Case



This data set was applied to the fusion processes described in Chapter 3. As stated for problem 5, the linear classifier was inadequate to get a reasonable classifier accuracy and the fused results suffered. Results were typical, but not reviewed here as the classifiers were poor.

4.9 Problem 2: 8 Feature Case RSM Study



The case study compared 5 factors in a full factorial experiment using IND_{LR} and NF_{ND} as described in Chapter 3. The 5 factors and their settings are displayed in Table 4-11.

Table 4-11: RSM Factor Settings.

Factor	*- (Low)*	*+ (High)*
A - C_{FP}	10	20
B - C_{FN}	5	9
C - C_{ND}	1	4
D - ρ	0	0.9
E - ρ_{auto}	0	0.9

The ISOC likelihood ratio heuristic was shown to be robust to correlation. The models found were not dependent on any levels of correlation. In fact, the main factor for all models tested was factor C, the cost of non-declaration. This factor proved to be 5 to 10 times more important than any other factor based on the sums of squares. Table 4-12 shows the factors for each model in terms of their importance by sums of squares. The models are generated for all possible indifference windows as described in Chapter 3.

Table 4-12: RSM Results for IND_{LR}.

delta 1	delta 2	1	2	3	4	Adj R^2	RMSE
		Factors and Interactions in Order of Importance					
0	**0**	C				0.9864	0.03
0	**0.1**	C	A			0.9857	0.03
0	**0.2**	C	A			0.9885	0.02
0	**0.3**	C	A			0.9880	0.03
0	**0.4**	C	A			0.9892	0.03
0	**0.5**	C	B			0.9857	0.03
0.1	**0**	C	B	A		0.9454	0.05
0.1	**0.1**	C	B	A		0.8752	0.07
0.1	**0.2**	C	B	A		0.8688	0.07
0.1	**0.3**	C	A	B		0.8529	0.08
0.1	**0.4**	C	B	A		0.8632	0.08
0.1	**0.5**	C				0.9988	0.01
0.2	**0**	C				0.9470	0.05
0.2	**0.1**	C	B	A		0.8753	0.07
0.2	**0.2**	C	B	A		0.8650	0.07
0.2	**0.3**	C	B	A		0.8511	0.08
0.2	**0.4**	C	B	A		0.8681	0.08
0.2	**0.5**	C				0.9948	0.02
0.3	**0**	C				0.9556	0.05
0.3	**0.1**	C	B	A		0.8790	0.07
0.3	**0.2**	C	B	A		0.8807	0.07
0.3	**0.3**	C	B	A		0.8741	0.07
0.3	**0.4**	C	B	A		0.8895	0.07
0.3	**0.5**	C	B	BC		0.9968	0.01
0.4	**0**	C				0.9415	0.06
0.4	**0.1**	C	A	B	AC	0.9166	0.06
0.4	**0.2**	C	B	A	AC	0.9145	0.06
0.4	**0.3**	C	B	A	BC	0.9000	0.06
0.4	**0.4**	C	B	BC		0.9133	0.07
0.4	**0.5**	C	B	BC		0.9982	0.01
0.5	**0**	C				1.0000	0.00
0.5	**0.1**	C				1.0000	0.00
0.5	**0.2**	C				1.0000	0.00
0.5	**0.3**	C				1.0000	0.00
0.5	**0.4**	C				0.9983	0.01
0.5	**0.5**	C				1.0000	0.00

The NF_{ND} technique was found to react to the correlation as shown in Table 4-13.

Table 4-13: NF_{ND} RSM Results.

	Factors and Interactions in order in Importance							
δ	**1**	**2**	**3**	**4**	**5**	**6**	**Adj R^2**	**RMSE**
0	A	D	E	B	DE		0.65	0.45
0.1	A	C	D	E	B	DE	0.59	0.41
0.2	C	A					0.56	0.55
0.3	C						0.70	0.57
0.4	C						0.86	0.48
0.5	C						1.00	0.00

Factors D (ρ) and E (ρ_{auto}) were found to be important in the first two models although as the indifference window size increased, the model became dependent on the cost of non-declaration only. The models also became more predictive as the window increased due to the associated increasing probability of non-declarations.

4.10 Problem 2: 8 Feature Case Likelihood Ratio Study 

This section attempted to prove IND_{LR} returned optimal rule sets as described in Chapter 3. Rule sets were generated and compared for problems 2 and 4 through complete enumeration of ISOC allowing non-declarations and IND_{LR}. This was accomplished for $\delta_j^i = (i\text{-}1) \times 0.05$ for $j = 1,2$ and $i = 1,2,\ldots,11$. IND_{LR} rules were then compared to the optimal rules found through complete enumeration of ISOC allowing non-declarations. These optimal rules always contained the IND_{LR} optimal rule showing that the likelihood ratio heuristic can reach an optimal rule set for the problems tested.

4.11 Summary

This chapter summarized the findings of this research. The heuristics described in Chapter 3 were compared to each other and neural networks. Costs and accuracies were compared.

5. Conclusions and Recommendations

5.1 Introduction

This chapter briefly reviews the research. The literature review, methodology and findings and analysis are each discussed. The important findings of this research process as a whole are reiterated and future research ideas are introduced.

5.2 Literature Review

Fratricide is unacceptable to the American public today. Air Force guidance supports this opinion by directing the use of fusion for improved accuracy and reliability of classifiers thus limiting the possibility of friendly fire incidents (AFPAM 14-410, 1998).

Two fusion methods considered in this research were the Identification System Operating Characteristic (ISOC) and Probabilistic Neural Network (PNN) fusion. The ISOC assumes information is independent while the PNN makes no such assumption. While the assumptions are clear, the effects due to dependence and correlation are not (Willett, *et al.*, 2000).

The data used for this experiment came from Leap (2004) as described in section 2.7. The data used in this research employed varied levels of correlation across 6 problems differing in geometry, mean vectors, and correlations ρ and ρ_{auto}.

5.3 Methodology

Indifference windows were introduced both at the individual classifier level as well as at the fused classifier level; ISOC methods utilized indifference windows at the individual classifier level while the neural networks allowed non-declarations at the fused level. The decision threshold was held constant at T = 0.5 to limit the number of

variables in this research; this seemed reasonable as the *a priori* class sizes were assumed equal. Three separate ISOC heuristics were developed for the addition of non-declarations both at the individual classifier level as well as at the fused indication level. Non-declaration heuristics were compared and contrasted through a cost function. IFD only allowed individual classifier non-declarations while IND_{PC} and IND_{LR} allowed fused non-declared indications. As a result, IFD non-declaration probabilities $P(ND_{IS})$ were constants for a given grid point (δ_1^i, δ_2^i). IND_{PC} and IND_{LR} non-declared probabilities were based on fused non-declared indications and varied based upon both the rule set used as well as the grid point location. A PNN fusion method was also developed to include non-declarations. This was used as a benchmark to compare the heuristics. A post-optimality analysis of the costs was performed to determine if there were significant interactions between costs and levels of correlation.

5.4 Findings

The heuristics developed above were comparable to the PNN fusion method. The ISOC methods continued to be robust to correlation with respect to accuracy and cost although costs were affected in some instances. ISOC fusion methods became more stable when sample size was sufficiently large; optimal individual classifier indifference window settings converged to a specific grid point location and optimal rule sets converged to a single set. The PNN fusion methods accuracies reacted appropriately with the introduction of correlation which was consistent with findings in Leap (2004). NF_{ND} and IND_{PC} were comparable methods of classification with similar costs, accuracies and non-declaration probabilities.

5.5 Implications

This research developed and tested heuristic approaches to finding the optimal Boolean fusion for a set of classifiers allowing non-declarations. The cost of non-declarations was found to be the overriding factor within the cost space as specified by ACC. Non-declaration schemes, such as those suggested in this research, have operating points which are less costly than forced decision methods but these solutions are sensitive to the cost of a non-declaration. Boolean fusion can find solutions which are comparable to those of feature level fusion methods when non-declarations are allowed. Feature level fusion requires a higher level of understanding than decision level fusion. An operator can understand individual sensor indications leading to a fused label while feature level fusion techniques such as neural networks require the testing of several parameters. These parameters can have a great effect on the fused indications out of the network. As a result, decision level fusion with non-declarations can be utilized to avoid inconsistencies while still reaching a satisfactory level of performance.

5.6 Future Research

This research was the next step in determining the effects of correlation on classifier fusion. One area for improvement to this study is the use of real world data for further confirmation. One further step would be to model ISOC fusion for dependent classifiers to consider more real world applications. Another improvement would be the consideration of a PNN compared with the ISOC heuristics described while allowing the PNN to train on an equivalent amount of data to see if the findings are consistent. Finally, the political correctness measure could be applied to the IND_{LR} as well as a difference measure to sort the rankings of states with no occurrences.

Bibliography

Air Force Doctrine Document 2-1, *Air Warfare*, Washington, D.C.: HQ USAF 2000.

Air Force Doctrine Document 2-5, *Information Operations*, Washington, D.C.: HQ USAF 1999.

Air Force Pamphlet 14-210, *USAF Intelligence Targeting Guide*, Washington, D.C.: HQ USAF, February 1998.

Alsing, Stephen G., Eric P. Blasch, and Kenneth W. Bauer Jr. "Three-Dimensional (3-D) Receiver Operating Characteristic (ROC) Trajectory Concepts for the Evaluation of Target Recognition Algorithms Faced with the Unknown Target Detection Problem." *Proc of SPIE Vol. 3718*, pp. 449-458, April 1999.

Chow, C. K, "On Optimum Recognition Error and Reject Tradeoff." *IEEE Transactions on Information Theory*. Vol. 16, No. 1:41-46. New York: IEEE Press, 1970.

Clemans, Paul P. "An Investigation of the Optimal Sensor Ensemble for Sensor Fusion." MS thesis, AFIT/GOR/ENS/04-03, Graduate School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2004.

Clutz, T., *A Framework for Prognostics Reasoning*. PhD Dissertation, AFIT/DS/ENS/03-01, Graduate School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2003.

Fukunaga, Keinosuke, and Raymond R. Hayes, "Effects of Sample Size in Classifier Design." *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Volume II, No. 8: pp. 873-885. New York: IEEE Press, August 1989.

Fumera, Giorgio, Fabio Roli, and Giorgio Giacinto. "Reject Option with Multiple Thresholds." *Pattern Recognition*, 33: 2099-2101, 2000.

Hall, David L. and Alan Steinberg. "Dirty Secrets in Multisensor Data Fusion.", Pennsylvania State University Applied Research Laboratory Tech. Report. 2001.

Haspert, J.K., "Optimum ID Sensor Fusion for Multiple Target Types." Institute for Defense Analysis (IDA) Document D2451, Virginia, March 2000.

Hebert, Adam J., "Building Battlespace Awareness," *Air Force Magazine Online*, vol 87: No 9 (September 2004). http://www.afa.org/magazine/sept2004/0904Isr.asp

Klein, Lawrence A. *Sensor and Data Fusion Concepts and Applications* (2^{nd} Edition), Washington: SPIE—The International Society for Optical Engineering, 1999.

Laine, Trevor, "Relevant Background for the Research of Sensor Fusion as Applied to Automatic Target Detection/Recognition (ATD/R) in an Environment with Correlated Input Data." *Advanced Applications for ANNs Special Study*, 2003.

-----, T. I. and Bauer, K.W., "A mathematical framework for the optimization of rejection and ROC thresholds in the fusion of correlated sensor data," Proc. of SPIE, Vol 5434, pp. 37-48, Orlando, FL April, 2004.

Leap, Nathan J. *An Investigation of the Effects of Correlation,Aautocorrelation, and Sample Size in Classifier Fusion*, MS thesis, AFIT/GOR/ENS/04-06, Graduate School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2004.

-----, Nathan J., Paul P. Clemans, Kenneth W. Bauer, Jr. and Mark E. Oxley, "An Investigation of the Effects of Correlation and Autocorrelation on Classifier Fusion and Optimal Classifier Ensembles", Proceedings of the Artificial Neural networks in Engineering Conference (ANNIE 2004), Editors: C. Dagli, D. Enke, A. Buczak, M. Embrechts, O. Ersoy, pp xxx-yyy, St. Louis, MO, 8-10 November 2004.

Oxley, M.E., and K. Bauer, "Classifier Fusion for Improved System Performance." *AFIT/ENS Working Document 02-02*, 2002.

Ralston, J.M., "Bayesian Sensor Fusion for Minimum-Cost I.D. Declaration." *1998 Joint Service Combat Identification Systems Conference on Requirements, Technologies and Developments,* (CISC-98), Volume 1 – Technical Proceedings. (1998)

-----, "Objective Rules of Engagement and Sensor Fusion in Combat Identification Systems." *Proceedings of the 1989 Symposium on Command and Control Research*, 3rd: Washington, DC, 1989.

Robinson, Guner S. and A. Omar Aboutalib, "Trade-Off Analysis of Multisensor Fusion Levels." *Proceedings of the 2nd National Symposium on Sensors and Sensor Fusion,* vol. II: 31 March 1989, pp. 21-34.

Roli, Fabio. Tutorial, AI*IA 2003 Tutorial *Fusion of Multiple Pattern Classifiers*, Dept. of Electrical and Electronics Eng., University of Cagliari, Italy, 2002.

Shipp, C.A and L.I. Kuncheva, "Relationships between combination methods and measures of diversity in combining classifiers." *Information Fusion*, vol 3, iss 2: 2 June 2002, pp. 135-138.

Specht, D. F, "Probabilistic Neural Networks." *Neural Networks: The official Journal of the International Neural Network Society*, vol 3: 109-118 (1990).

-----, “Probabilistic Neural Networks and the Polynomial Adaline as Complementary Techniques for Classification.” *IEEE Transactions on Neural Networks*. Vol. 1, No.1 pp 111-121. New York: IEEE Press, March 1990.

Storm, S. A., K. Bauer, and M. Oxley, “An Investigation in the Effects of Correlation in Classifier Fusion.” *Intelligent Engineering Systems Through Artificial Neural Networks*, vol 13: 2003, pp.619-624.

-----, Susan A., Kenneth W. Bauer, Jr. and Mark E. Oxley “An Investigation of the Effects of Correlation in Classifier Fusion”, Proceedings of the Artificial Neural networks in Engineering Conference (ANNIE 2003), Editors: C. Dagli, A. Buczak, J, Ghosh, M. Embrechts, O. Ersoy, pp 619-624, 2003

United States General Accounting Office, *Combat Identification Systems: Strengthened Management Efforts to Ensure Required Capabilities*. Report to the Secretary of Defense, GAO-01-632; U.S Government Printing Office, June 2001.

Waltz, Edward and Llinas, James. *Multisensor Data Fusion*, Norwood: Artech House, INC., 1990.

Wasserman, P., and V. Nostrand, *Advanced Methods in Neural Computing.* 1993.

Willet, P., P.F. Swaszek, and R.S. Blud, “The good, bad and ugly: distributed detection of a known signal in dependent Gaussian noise.” *IEEE Transactions on Signal Processing*, vol 48, iss 12:pp. 3266-3279, 12 December 2000.

Vita

Captain Frank M Mindrup graduated from Ponderosa High School in Parker, Colorado. He entered undergraduate studies at the United States Air Force Academy (USAFA) where he graduated with a Bachelor of Science Degree in Operations Research in June 1999. He was commissioned through USAFA with a Reserve Commission.

His first assignment was at USAFA Preparatory School as an instructor of physics and chemistry and football coach. He was then assigned to LAAFB, El Segundo, CA as a financial management officer for the Titan Launch Programs Systems Program Office. In August 2004, he entered the Graduate School of Engineering and Management, Air Force Institute of Technology. Upon graduation, he will be assigned to Air Force Space Command, Schriever AFB, CO.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 074-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to an penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
16-03-2005	**Master's Thesis**	Aug 2003 – Mar 2005

4. TITLE AND SUBTITLE	
AN INVESTIGATION OF THE EFFECTS OF CORRELATION AND AUTOCORRELATION IN CLASSIFIER FUSION WITH NON-DECLARATIONS	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
Mindrup, Frank M., Capt., USAF	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)	8. PERFORMING ORGANIZATION REPORT NUMBER
Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way, Building 641 WPAFB OH 45433-7765	AFIT/GOR/ENS/05-13

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
ATTN: Neal Glassman AFOSR/NM 4015 Wilson Blvd Mail Room 713 Arlington, VA 22203-1954	ATTN: Charles Sadowski ACC/DRSA 216 Hunting Ave, Room 106 Langley AFB, Va 23665-2777	AFOSR/NM, ACC/DRSA
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Air Force doctrine requires reliable and accurate information when striking targets. Further, this doctrine states that fusion should be utilized whenever possible to ensure the best possible information is conveyed; there is no specific guidance as to how to fuse this information. This thesis extends the research found in Leap, Bauer, and Oxley (2004) to include a non-declared class. The Identification system operating characteristic (ISOC) was adapted to allow for non-declarations both at the individual sensor level as well as the fused output level. A probabilistic neural network (PNN) was also used as a fusion technique. A cost function was developed that incorporated misclassification error as well as non-declaration rules. In addition, a heuristic was developed to find optimal rules through a likelihood ratio method. Finally, a sensitivity analysis was performed.

15. SUBJECT TERMS

Sensor Fusion, Classifier Fusion, Classification, Probabilistic Neural Network, ISOC Fusion, Correlation, Autocorrelation, Sample Size

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Kenneth W. Bauer, AFIT/ENS
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	104	(937) 255-6565, ext 4328; e-mail: Kenneth.Bauer @afit.edu

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18